

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 10, 2004, 15:29:15 ; Search time 372 Seconds
(without alignments)
902.171 Million cell updates/sec

Title: US-09-972-032-2
Perfect score: 456
Sequence: 1 MCGRPVRVSAGCFADAHWT.....SAGLTVRDRPQLGELCMGRG 79

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US0972032/runat_06072004_121451_8379/app_query.fasta_1.263
-DB=N_Geneseq_29Jan04 -Qfmt=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNIT=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USPR=US0972032 @CIGN 1.1 470 @runat_06072004_121451_8379 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04:.*
1: Geneseq1980s:.*
2: Geneseq1980s:.*
3: Geneseq2000s:.*
4: Geneseq2001as:.*
5: Geneseq2001bs:.*
6: Geneseq2002s:.*
7: Geneseq2003as:.*
8: Geneseq2003bs:.*
9: Geneseq2003cs:.*
10: Geneseq2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	990	6	ABL60606 Human ERC
C 2	164.5	36.1	597	5	AAS69040 DNA encod
C 3	95	20.8	349901	9	ADC86940 Human GPC
4	94.5	20.7	24081	7	AAD54223 Streptomy
5	94.5	20.7	52101	7	AAD54217 Streptomy
C 6	94.5	20.6	696	3	AAS50885 A. sulcat
7	93.5	20.5	606	5	AAS52281 Human AFP
8	93.5	20.5	943	6	ABK70016 cDNA enco

9	93.5	20.5	943	8	ADA01379 Human PRO
10	93.5	20.5	943	8	ADA43808 Human cDN
11	93.5	20.5	943	8	ADA43576 Human cDN
12	93.5	20.5	943	8	ADA01251 Human PRO
13	93.5	20.5	943	8	ADA01135 Human cDN
14	93.5	20.5	943	8	ADA43692 Human cDN
15	93.5	20.5	943	8	ADA06954 Human PRO
16	93.5	20.5	943	8	ADA08442 Novel hum
17	93.5	20.5	943	8	ADB99735 Human PRO
18	93.5	20.5	943	8	ADB87018 Human PRO
19	93.5	20.5	943	8	ADB66173 Human cDN
20	93.5	20.5	943	9	ADB99851 Human PRO
21	93.5	20.5	943	9	ADB99506 Novel hum
22	93.5	20.5	943	9	ADB66057 Human cDN
23	93.5	20.5	943	9	ADC23455 Human cDN
24	93.5	20.5	943	9	ADC26148 Human PRO
25	93.5	20.5	943	9	ADE04975 Human PRO
26	93.5	20.5	943	9	ADE11281 Human PRO
27	93.5	20.5	943	9	ADD88212 Human PRO
28	93.5	20.5	943	9	ADD95507 Human cDN
29	93.5	20.5	943	9	ADE06437 Human PRO
30	93.5	20.5	943	9	ADD88328 Human PRO
31	93.5	20.5	943	9	ADD90909 Human cDN
32	93.5	20.5	943	10	ADE51762 Human cDN
33	93.5	20.5	943	10	ADE51878 Human cDN
34	93.5	20.5	943	10	ADE37736 Human cDN
35	93.5	20.5	943	10	ADE37620 Human cDN
36	93.5	20.5	943	10	ADD95391 Human cDN
37	93.5	20.5	943	10	ADE38091 Human PRO
38	93.5	20.5	943	10	ADE76180 Human PRO
39	93.5	20.5	943	10	ADE39503 Human PRO
40	93.5	20.5	943	10	ADE04307 Human PRO
41	93.5	20.5	943	10	ADE39904 Human PRO
42	93.5	20.5	943	10	ADE19769 Human PRO
43	93.5	20.5	943	10	ADE77347 Human cDN
44	93.5	20.5	943	10	ADE65455 Human PRO
45	93.5	20.5	943	10	ADE65455 Human PRO

ALIGNMENTS

RESULT 1
ABL60606
ID ABL60606 standard; cDNA; 990 BP.
XX
AC ABL60606;
XX
DT 27-AUG-2002 (first entry)
XX
DE Human ERCoA3 protein encoding cDNA.
XX
KW Estrogen Receptor Coregulator 3; ERCoA3; tamoxifen; estrogen; cancer;
XX osteoporosis; cytostatic; osteopathic; human; gene; ss.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 203..442
FT /*tag= a
FT /product= "ERCoA3 protein"

PN WO200228352-A2.
XX
PD 11-APR-2002.
XX
PF 05-OCT-2001; 2001WO-US031271.
XX
PR 05-OCT-2000; 2000US-0238190P.
XX
PA (UYCA-) UNIV CASE WESTERN RESERVE.
XX
PI Montano M, Sutton A;
XX

DR WPI; 2002-454492/48.
 DR P-PSDB; ABB08035.
 XX
 PT New polypeptide, that is a functional equivalent of ERCoA3 (Estrogen
 PT Receptor Coregulator 3), is useful in inhibiting or reducing tamoxifen or
 PT estrogen-induced proliferation of cancer cells and in treating
 PT osteoporosis.
 XX
 PS Claim 5; Fig 1; 39pp; English.
 XX
 CC The invention relates to a ERCoA3 (Estrogen Receptor Coregulator 3)
 CC protein and encoding polynucleotides. ERCoA3 can be used to inhibit or
 CC reduce tamoxifen or estrogen induced proliferation of cancer cells, by
 CC reduced activity of ERCoA3, and for detecting cancer cells that are
 CC tamoxifen resistant, or to treat osteoporosis, by increasing levels of
 CC ERCoA3 in cells. The encoding polynucleotide can be used to inhibit
 CC translation of a mRNA encoding ERCoA3. ERCoA3 acts as a coregulator
 CC protein and can bind to the estrogen receptor to activate a molecular or
 CC cellular response in the cell. The present sequence represents the human
 CC ERCoA3 encoding cDNA
 XX
 SQ Sequence 990 BP; 189 A; 281 C; 317 G; 203 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.2e-26 Length: 990
 Score: 456.00 Matches: 79
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-09-972-032-2 (1-79) x ABL60606 (1-990)
 QY 1 MetCysGlyArgProArgArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThr 20
 DB 203 ATGTGTGGAGACCGCGTCGTAAGCGCTGGATGGCTTCGCTGATGCACATTGGACC 262
 QY 21 GlyLeuTrpThrGlyLeuGlyGlnGlyGlyLeuGlyGlyLeuGlyGlnAla 40
 DB 263 GGGCTCTGGACTGGCTAGCGGAAAGGCGGAAATGGGCGCCGAGGCGCAGGCC 322
 QY 41 SerProThrProAspCysAlaSerArgTrpProArgSerAlaSerArgTrpProTrpSer 60
 DB 323 TCGCCGACCCCGACTCGGCTCCGGTGGCCCGCAGCGCTCCCGGTGGCCCTGGACT 382
 QY 61 AlaGlyLeuThrValArgAspArgProGlnLeuGlyGlyLeuCysMetGlyArgGly 79
 DB 383 GCAGGTCTTACCGTCCGAGATCGTCGCAACTGGCGAGCTGTGCATGGCGGTGCG 439
 RESULT 2
 ID AAS69040 standard; cDNA; 597 BP.
 XX
 AC AAS69040;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #4844.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 EN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX
 AC AAS69040;

(HYSE-) HYSEQ INC.
 Drmanac RT, Liu C, Tang YT;
 WPI; 2001-639362/73.
 P-PSDB; ABB04853.
 New isolated polynucleotide and encoded polypeptides, useful in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits and to assess
 biodiversity.
 Claim 1; SEQ ID NO 4844; 103pp; English.
 The invention relates to isolated polynucleotide (I) and polypeptide (II)
 sequences. (I) is useful as hybridisation probes, polymerase chain
 reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 and in recombinant production of (II). The polynucleotides are also used
 in diagnostics as expressed sequence tags for identifying expressed
 genes. (I) is useful in gene therapy techniques to restore normal
 activity of (II) or to treat disease states involving (II). (II) is
 useful for generating antibodies against it, detecting or quantitating a
 polypeptide in tissue, as molecular weight markers and as a food
 supplement. (II) and its binding partners are useful in medical imaging
 of sites expressing (II). (I) and (II) are useful for treating disorders
 involving aberrant protein expression or biological activities. The
 polypeptide and polynucleotide sequences have applications in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits to assess biodiversity
 and to produce other types of data and products dependent on DNA and
 amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 coding sequences of the invention. Note: The sequence data for this
 patent did not appear in the printed specification, but was obtained in
 electronic format directly from WIPO at
 ftp.wipo.int/pub/published_pct_sequences
 Sequence 597 BP; 97 A; 217 C; 196 G; 87 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.000237 Length: 597
 Score: 164.50 Matches: 43
 Percent Similarity: 53.33% Conservative: 5
 Best Local Similarity: 47.78% Mismatches: 26
 Query Match: 36.07% Indels: 17
 DB: 5 Gaps: 3
 US-09-972-032-2 (1-79) x AAS69040 (1-597)
 QY 1 MetCysGlyArgProArgArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThr 20
 DB 522 ATGTGTGGAGACCGCGTCGCGTAAGCGCTGGATGGCTTCGCTGATGCACATTGGACC 463
 QY 21 GlyLeuTrpThrGlyLeuGlyGlnGlyGlyLeuGlyGlyLeuGlyGlnAla 40
 DB 462 GGGCTCTGGACTGGCT-GGCAGGTGTCCCTGCGCGCCCTCTGGGTAGCCAGCACAGCA 404
 QY 41 Ser-----ProThrProAspCys-----AlaSer 48
 DB 403 TCCCTGGCAATATCTGGAGAGTAGGTTGAAGGAGCGGGCTGCTGTAGACATTGAC 344
 QY 49 ArgTrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThrValArgAspArg 68
 DB 343 AGGTGGCCTCGCCAGCAGCGCTTAGCTCCA-----CTTCCTGTAGCACCAGG 296
 QY 69 ProGlnLeuGlyGlnLeuCysMetGlyArg 78
 DB 295 GGAGAGGCGAGAGGATCTGCCTCTGGACGT 266
 RESULT 3
 ID ADC86940/c
 XX
 AC ADC86940 standard; DNA; 349901 BP.
 XX
 AC ADC86940;

```

XX 01-JAN-2004 (first entry)
XX Human GPCR gene SEQ ID NO:1393.
XX ds; gene; human; GPCR;
XX guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX Homo sapiens.
XX EP1270724-A2.
XX 02-JAN-2003.
XX 18-JUN-2002; 2002EP-00013517.
XX 18-JUN-2001; 2001JP-00246789.
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX Suwa M, Asei K, Akiyama Y, Aburatani H;
XX WPI; 2003-315783/31.
XX P-PSDB; ADC86941.
XX New polynucleotide, useful for preparing a composition for treating a
XX patient in need of increased or suppressed activity or expression of the
XX guanosine triphosphate-binding protein coupled receptor.
XX Claim 1; SEQ ID NO 1393; 28pp; English.
XX The invention relates to a novel polynucleotide encoding a guanosine
XX triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
XX the invention may have a use in gene therapy. The polynucleotide and
XX polypeptide are useful for preparing a composition for treating a patient
XX in need of increased or suppressed activity or expression of the
XX guanosine triphosphate-binding protein coupled receptor. The
XX polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
XX invention.
XX
XX SQ Sequence 349901 BP; 106873 A; 66040 C; 67419 G; 108968 T; 0 U; 601 Other;
Alignment Scores:
Pred. No.: 3.18e+04 Length: 349901
Score: 95.00 Matches: 37
Percent Similarity: 37.39% Conservative: 6
Best Local Similarity: 32.17% Mismatches: 29
Query Match: 20.83% Indels: 43
DB: 9 Gaps: 5
US-09-972-032-2 (1-79) x ADC86940 (1-349901)
QY 3 GlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeu 22
Db 107177 GCGGGCACTGTCGGGAGACTGGGCTGC-----GCAGGAGCCACGGCAGTGG--- 107127
QY 23 TrpThrGlyLeuGlyGlnGlnGly----- 32
Db 107126 -----GGAGGAGGCTTAGGCATGCGGGCTGTGTGTCCACGCCCTACAGGGAA 107073
QY 32 ----- 32
Db 107072 GCAGCTAAGCCCTGTGAGAAATCGAGACAGAGCTGTGGCCCGAGTCTAAGCCCT 107013
QY 33 ---GlyIleGlyProGlnGlyGlnAlaSerProThrProAspCysAlaSerArgTrpPro 51
Db 107012 CACTGCTGGGGCAGCGGGCCCGCGAGCCACCGCCACAGAACTCGCGTGGCC 106953
QY 52 ArgSerAlaSerArgTrpProTrpSerAla---GlyLeuThrValArgAspArgProGln 70
Db 106952 ACAAGCACCGCGGCGAGCCCTTGGGTCTCTGCCACACCTCTCCCTCCACACCTCCCGCAA 106893

```

```

QY 71 -LeuGlyCluLeuCysMet-----GlyArgGly 79
Db 106892 GCTGAGGAGCTGCTCTAGCCTCGCCAGCCCGAGAGGGC 106850
RESULT 4
AAD54223
ID AAD54223 standard; DNA; 24081 BP.
XX AC AAD54223;
XX 17-JUN-2003 (first entry)
XX Streptomyces platensis subspecies rosaceus dorrigocin ORF6 DNA.
XX Polyketide biosynthesis; dorrigocin; DORR; lactimidomycin; LACT; gene;
XX ds.
XX Streptomyces platensis.
XX Key Location/Qualifiers
XX CDS 1..24081
XX FT /*tag= a
XX FT /product= "ORF6 protein"
XX FT /note= "No start codon"
XX FT /partial
XX WO200288176-A2.
XX 07-NOV-2002.
XX 26-APR-2002; 2002WO-CA000591.
XX 26-APR-2001; 2001US-0286346P.
XX (ECOP-) ECOPIA BIOSCIENCES INC.
XX Farnet CM, Zazopoulos E, Staffa A, Yang X;
XX WPI; 2003-201222/19.
XX DR P-PSDB; AAE35489.
XX
XX Novel isolated or purified polypeptide involved in biosynthesis of
XX polyketide dorrigocin or polyketide lactimidomycin, useful for preparing
XX dorrigocin or lactimidomycin.
XX Claim 6; Page 168-181; 312pp; English.
XX The invention relates to novel proteins involved in the biosynthesis of
XX polyketide dorrigocin (DORR) or lactimidomycin (LACT) biosynthesis by
XX microorganisms. Sequences of the invention allow direct manipulation of
XX dorrigocin, lactimidomycin and related chemical structures via chemical
XX engineering of the enzymes involved in the biosynthesis of dorrigocin and
XX lactimidomycin. They are useful for introducing chemical handles into
XX normally inert positions that permit subsequence chemical modifications
XX and facilitate the development of polyketides. The genes and proteins of
XX the invention can also be used to generate a focused library of analogues
XX around a polyketide lead candidate to fine-tune the compound for optimal
XX properties. They are useful for generating antibodies specific for the
XX polyketide biosynthesis. The present sequence is Streptomyces platensis
XX subspecies rosaceus DORR ORF6 DNA
XX
XX SQ Sequence 24081 BP; 3269 A; 9645 C; 8186 G; 2981 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 2.41e+03 Length: 24081
Score: 94.50 Matches: 27
Percent Similarity: 42.42% Conservative: 1
Best Local Similarity: 40.91% Mismatches: 30
Query Match: 20.72% Indels: 8
DB: 7 Gaps: 3
US-09-972-032-2 (1-79) x AAD54223 (1-24081)

```

QY 9 SerAlaGlyCysGlyPheAlaSerPheAlaHisTrpThrGlyLeuTrpThrGlyLeuGly-Gl 28
 Db 12490 TCGGGCGCTGGCGCTTCGAGACGGCAGCGCGCGCGCTCGCGAGCGCTGGACTGC 12549
 QY 28 uGlyGlnGluGlyGlyIleGlyProGluGlyGlnAlaSerProThrProAspCysAlaase 48
 Db 12550 CGGGTCAAGTGTCTCAACTGGGCTACTGGGCAACGTCCTCGAGCTCTCGCGAC 12609
 QY 48 r-----ArgTrpProArgSer-----AlaSerArgTrpPro---TrpSerAl 61
 Db 12610 GTGACGTCCATGGGACTGGCGCGCGATCGCGCGCAACGGCGATGGCGCGACTGGAGCGC 12669
 QY 61 aGlyLeuThrValArg 66
 Db 12670 CTCCTGGCGCGCGCGC 12685
 RESULT 5
 AAD54217
 ID AAD54217 standard; DNA; 52101 BP.
 AC AAD54217;
 DT 17-JUN-2003 (first entry)
 DE Streptomyces platensis subspecies rosaceus dorrigin DNA.
 XX KW Polyketide biosynthesis; dorrigin; DORR; lactimidomycin; LACT; gene;
 XX OS ds.
 XX ST Streptomyces platensis.
 FH Key Location/Qualifiers
 CDS complement(67..3720)
 FT /*tag= a
 FT /product= "ORF1 protein"
 FT 4092..5681
 FT /*tag= b
 FT /product= "ORF2 protein"
 FT /note= "No start codon"
 FT /partial
 FT 5767..6018
 FT /*tag= c
 FT /product= "ORF3 protein"
 FT 6023..7993
 FT /*tag= d
 FT /product= "ORF4 protein"
 FT 8009..17587
 FT /*tag= e
 FT /product= "ORF5 protein"
 FT 17634..41714
 FT /*tag= f
 FT /product= "ORF6 protein"
 FT /note= "No start codon"
 FT /partial
 FT 41772..47633
 FT /*tag= g
 FT /product= "ORF7 protein"
 FT /note= "No start codon"
 FT /partial
 FT 47635..49890
 FT /*tag= h
 FT /product= "ORF8 protein"
 FT 49922..50938
 FT /*tag= i
 FT /product= "ORF9 protein"
 FT 51234..52079
 FT /*tag= j
 FT /product= "ORF10 protein"
 XX HQ200288176-A2.
 XX PD 07-NOV-2002.

PF 26-APR-2002; 2002WO-CA000591.
 XX
 PR 26-APR-2001; 2001US-0286346P.
 XX
 PA (ECOP-) ECOPIA BIOSCIENCES INC.
 XX
 PI Farnet CM, Zazopoulos E, Staffa A, Yang X;
 XX
 DR WPI: 2003-201222/19
 DR P-FSDS; AAE35486, AAE35485, AAE35486, AAE35487, AAE35488, AAE35489,
 DR AAE35490, AAE35491, AAE35492, AAE35493.
 XX
 PT Novel isolated or purified polypeptide involved in biosynthesis of
 PT polyketide dorrigin or polyketide lactimidomycin, useful for preparing
 PT dorrigin or lactimidomycin.
 XX
 PS Claim 1; Page 85-113; 312pp; English.
 XX
 CC The invention relates to novel proteins involved in the biosynthesis of
 CC polyketide dorrigin (DORR) or lactimidomycin (LACT) biosynthesis by
 CC microorganisms. Sequences of the invention allow direct manipulation of
 CC dorrigin, lactimidomycin and related chemical structures via chemical
 CC engineering of the enzymes involved in the biosynthesis of dorrigin and
 CC lactimidomycin. They are useful for introducing chemical handles into
 CC normally inert positions that permit subsequent chemical modifications
 CC and facilitate the development of polyketides. The genes and proteins of
 CC the invention can also be used to generate a focused library of analogues
 CC around a polyketide lead candidate to fine-tune the compound for optimal
 CC properties. They are useful for generating antibodies specific for the
 CC polyketide biosynthesis. The present sequence is Streptomyces platensis
 CC subspecies rosaceus DORR DNA
 XX
 SQ Sequence 52101 BP; 7364 A; 20113 C; 17894 G; 6730 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 5.21e+03 Length: 52101
 Score: 94.50 Matches: 27
 Percent Similarity: 42.42% Conservative: 1
 Best Local Similarity: 40.91% Mismatches: 30
 Query Match: 20.72% Indels: 8
 DB: 7 Gaps: 3
 US-09-972-032-2 (1-79) x AAD54217 (1-52101)
 QY 9 SerAlaGlyCysGlyPheAlaSerPheAlaHisTrpThrGlyLeuTrpThrGlyLeuGly-Gl 28
 Db 30123 TCGGGCGCTGGCGCTTCGAGACGGCAGCGCGCGCGCTCGCGAGCGCTGGACTGC 30182
 QY 28 uGlyGlnGluGlyGlyIleGlyProGluGlyGlnAlaSerProThrProAspCysAlaase 48
 Db 30183 CGGGTCAAGTGTCTCAACTGGGCTACTGGGCAACGTCCTCGAGCTCTCGCGAC 30242
 QY 48 r-----ArgTrpProArgSer-----AlaSerArgTrpPro---TrpSerAl 61
 Db 30243 GTGACGTCCATGGGACTGGCGCGCGATCGCGCGCAACGGCGATGGCGCGACTGGAGCGC 30302
 QY 61 aGlyLeuThrValArg 66
 Db 30303 CTCCTGGCGCGCGCGC 30318
 RESULT 6
 AAA50885/c
 ID AAA50885 standard; cDNA; 696 BP.
 XX
 AC AAA50885;
 XX
 DT 08-SEP-2000 (first entry)
 XX
 DE A. sulcata humanised mutant fluorescent protein, Mut1, coding sequence.
 XX KW Green fluorescent protein; GFP; Mut1; protein localisation;
 XX fluorescence resonance energy transfer; mutant; ss.
 XX

Db	326	CGCTCACCGCAGCCAGCTCACCCTCTCTGGGGCTGTTGGGCCTGCTGCTGCTAC	385
Qy	66	gaspargProGlnLeu-----GlyIleuCys	76
Db	386	CCGACGCCCGCTGCTGGGAGGAGGCATGGCCGCTGCATCCCACTGGCGAGTTTGTCC	445
Qy	76	tgly 77	
Db	446	TGGT 449	
RESULT 8			
ABK70016			
ID	ABK70016	standard; DNA; 943 BP.	
XX	AC		
XX	ABK70016;		
DT	15-JUL-2002	(first entry)	
DE	cDNA encoding human Pro peptide #56.		
XX	Human; ss; gene; PRO; secreted protein; transmembrane protein;		
XX	genetic disorder; tumour; cancer.		
XX	Homo sapiens.		
OS			
XX	WO200224888-A2.		
XX	28-MAR-2002.		
PD			
XX			
PF	29-AUG-2001; 2001WO-US027099.		
XX			
PR	01-SEP-2000; 2000US-0229896P.		
PR	05-SEP-2000; 2000US-0230621P.		
PR	22-SEP-2000; 2000US-0235147P.		
PR	10-NOV-2000; 2000WO-US030873.		
PR	12-JAN-2001; 2001US-0261878P.		
PR	16-JAN-2001; 2001US-0261910P.		
PR	16-JAN-2001; 2001US-0261939P.		
PR	16-JAN-2001; 2001US-0262150P.		
PR	25-JAN-2001; 2001US-0264395P.		
PR	02-FEB-2001; 2001US-0266421P.		
PR	09-FEB-2001; 2001US-0267623P.		
PR	28-FEB-2001; 2001WO-US006520.		
PR	09-MAR-2001; 2001US-0274399P.		
PR	03-APR-2001; 2001US-0280982P.		
PR	04-APR-2001; 2001US-0282129P.		
PR	04-APR-2001; 2001US-0282199P.		
PR	09-MAY-2001; 2001US-0290589P.		
PR	25-MAY-2001; 2001WO-US017092.		
PR	01-JUN-2001; 2001WO-US017800.		
PR	20-JUN-2001; 2001WO-US019692.		
PR	29-JUN-2001; 2001WO-US021086.		
PR	09-JUL-2001; 2001WO-US021735.		
XX			
XX	(GETH) GENENTECH INC.		
XX			
PI	Baker KP, Eaton DL, Filvaroff E, Goddard A, Grimaldi JC;		
PI	Gurney AL, Smith V, Stephan J, Watanabe CK, Wood WI, Zhang Z;		
PI	Fong S;		
XX			
DR	WPI; 2002-362426/39.		
DR	P-PSDB; ABG34085.		

CC recombinant means and antibodies specific for the protein of the
CC invention. The antibody may be used for detecting the PRO proteins of the
CC CC invention and may be used to modify their activity. polynucleotides may
CC be used as hybridisation probes for a cDNA library to isolate the full-
CC length PRO cDNA or to isolate other cDNAs, to construct hybridisation
CC probes for mapping the gene which encodes that PRO and for genetic
CC analysis of individuals with genetic disorders, in assays to identify
CC other proteins or molecules involved in binding reaction, to generate
CC transgenic animals or knock-out animals which in turn are useful in the
CC development and screening of therapeutically useful reagents, for
CC chromosome identification, and tissue typing. The PRO polypeptides are
CC useful in gene therapy, and as molecular weight markers for protein
CC electrophoresis purposes. The sequences may also be used to detect
CC overexpression on PRO polypeptides in cancerous tumours and for screening
CC for differentially expressed genes using microarray technology. The
CC present sequence represents a cDNA encoding a human PRO protein of the
CC invention
XX

SQ Sequence 943 BP; 152 A; 306 C; 299 G; 186 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 114 Length: 943
Score: 93.50 Matches: 36
Percent Similarity: 35.25% Conservative: 7
Best Local Similarity: 29.51% Mismatches: 31
Query Match: 20.50% Indels: 48
DB: 6 Gaps: 6

US-09-972-032-2 (1-79) x ABK70016 (1-943)

QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
Db 138 TGTGCCAGACGCTGGAGGATGGGGCGCAGGCCGACGCTGGGGCTGTGGAGGTCTCTGTGTGC 197
QY 22 LeuTrpThrGlyLeuGlyGluGlyGlnGlyGlyIleGlyProGlu----- 37
Db 198 TGTGTGGACAGACCCGGGAGGCCGA-----GCCCTGGGGCCAGAGCCGCCAGGTGG 251
QY 38 -----GlyGlnAlaSerProThrPro 44
Db 252 AGCCACATGACTGTGAGGCGCTGGGCTCCGAGGCAGCGCGCTCCAGGAGTCCC 311
QY 45 Asp-----CysAlaSerArg-----TrpProArg----- 52
Db 312 GAGGCACCGCTCAAACTGCAGTTCGACATGATGCGCGCTGCACACCTGGTGGCCACGCGCG 371
QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66
Db 372 CGCTCAGCGCAGGCCAGCTACCTTCTCTGGGGCTGGTGGGCTTGGCCCTGCTGTCTAC 431
QY 66 GasPArgProGlnLeu-----GlyGlnLeuCysMe 76
Db 432 CCGAGCCCGCGTGTGGGAGGAGGCCATGCGCGCTGCATTCCTCAACTGGCGAGTTTGTGCC 491
QY 76 tgly 77
Db 492 TGGT 495

RESULT 9
ADA01379
ID ADA01379 standard; cDNA; 943 BP.
XX
XX ADA01379;
AC
XX
XX 06-NOV-2003 (first entry)
DT
XX
XX Human PRO polynucleotide #56.
DE
XX Human; gene; ss; PRO; secreted polypeptide; transmembrane polypeptide;
XX tumour necrosis factor-alpha; TNF-alpha; blood; chondrocyte cell; tumour;
KW adrenal; lung; colon; breast; prostate; rectum; cervix; liver; cancer;
KW microvascular endothelial cell; endothelial cell tube formation;
KW sports-related joint problem; articular cartilage defect; osteoarthritis;
KW

CC an oligonucleotide probe derived from any one of the above nucleotide
 CC sequences. PRO6018 polypeptide is useful for stimulating the
 CC proliferation or differentiation of chondrocyte cells. PRO1313, PRO20080
 CC and PRO21383 polypeptides are useful for stimulating the proliferation of
 CC human microvascular endothelial cells. PRO6071, PRO4487 and PRO6006
 CC polypeptides are useful for inhibiting the proliferation of human
 CC microvascular endothelial cells. PRO polypeptides are useful for
 CC detecting the presence of tumour in a mammal, including tumours of lung,
 CC colon, breast, prostate, rectal, kidney and liver. PRO281, PRO1560,
 CC PRO189, PRO4499, PRO6308, PRO6000, PRO10275, PRO21207, PRO20933 and
 CC PRO34274 polypeptides are useful for inducing endothelial cell tube
 CC formation. PRO or the antibody are useful in the preparation of a
 CC medicament for treating a condition responsive to PRO polypeptide. The
 CC oligonucleotide probes are useful for isolating genomic and cDNA
 CC nucleotide sequences, for measuring or detecting the expression of an
 CC associated gene, and as antisense probes. PRO nucleic acid is useful as a
 CC hybridisation probe, in chromosome and gene mapping, in the generation of
 CC antisense RNA and DNA, and for the preparation PRO polypeptides. PRO and
 CC PRO nucleic acid are useful as therapeutic agents, e.g. vaccines. The
 CC present sequence encodes a PRO protein.
 XX
 SQ Sequence 943 BP; 152 A; 306 C; 299 G; 186 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 114 Length: 943
 Score: 93.50 Matches: 36
 Percent Similarity: 35.25% Conservative: 7
 Best Local Similarity: 29.51% Mismatches: 31
 Query Match: 20.50% Indels: 48
 DB: 8 Gaps: 6

US-09-972-032-2 (1-79) x ADA43808 (1-943)

QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
 Db 138 TGTGCCAGACCTGGAGGATGGCGGAGCGGAGCGGAGCTGGGCTGTGGAGTCTGCTGGC 197
 QY 22 LeuTrpThrGlyLeuGlyGluGlyGlnGlyGlyGlyGlyGlyGlyGlyGlyGly 37
 Db 198 TGGTGGACAGACCGCGGAGGCGCGA-----GCCCTGGGCGCAGAGCGCGCAGGTGG 251
 QY 38 -----GlyGlnAlaSerProThrPro 44
 Db 252 ACGCACATGACTGTGAGCGCTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGTGGGT 311
 QY 45 Asp-----CysAlaSerArg-----TrpProArg----- 52
 Db 312 GAGGACCGGTCAAACTGCAGTTCGACATGATGGCGCTGCAACCTGTGGCCAGCGCG 371
 QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66
 Db 372 CGCTACCGCAGCCAGCTACCTTCTCTCTGGGCTGTGGGCTGTGGGCTGTGGGCTGT 431
 QY 66 GAspArgProGlnLeu-----GlyGluLeuCysMe 76
 Db 432 CCAGCGCCCGCTGCTGGAGGAGGCATGCGCGCTGCATTCACATGGCGAGTTTGTC 491
 QY 76 tGly 77
 Db 492 TGGT 495

RESULT 11

ID ADA43576 standard; cDNA; 943 BP.

AC ADA43576;

DT 20-NOV-2003 (first entry)

XX Human cDNA encoding secreted/transmembrane polypeptide PRO21434.

DE ss; gene; human; PRO; secreted protein; transmembrane protein;

KW endothelial cell tube formation; chondrocyte cell differentiation;

KW microvascular endothelial cell; tumour; lung tumour; colon tumour;
 KW breast tumour; prostate tumour; rectal tumour; kidney tumour;
 XX liver tumour; cytostatic; vaccine.
 OS Homo sapiens.
 XX US2003073196-A1.
 XX PD 17-APR-2003.
 XX 18-SEP-2002; 2002US-00246210.
 PP 04-APR-2001; 2001US-0282199P.
 PR 29-AUG-2001; 2001WO-US027099.
 PR 18-JUL-2002; 2002US-00197942.
 XX (GETH) GENENTECH INC.
 PA Baker KP, Eaton DL, Filvaroff E, Goddard A, Grimaldi JC;
 PI Gurney AL, Smith V, Stephan JP, Watanabe CK, Wood WI, Zhang Z;
 PI Fong S;
 XX WPI; 2003-743814/70.
 DR P-PSDB; ADA43577.
 XX
 PT New secreted and transmembrane PRO polypeptides, e.g. PRO20080 or
 PT PRO21383 useful for stimulating the proliferation or differentiation of
 PT chondrocyte cells and detecting the presence of a tumor in a mammal.
 XX
 PS Claim 2; Fig 111; 307pp; English.
 XX

CC The invention relates to an isolated secreted/transmembrane (PRO)
 CC polypeptide, having at least 80% sequence identity to a sequence selected
 CC from any one of the 57 amino acid sequences given in specification, or to
 CC a sequence encoded by a nucleic acid molecule selected from any one of
 CC the nucleic acids deposited under any of the ATCC accession numbers given
 CC in specification, or a sequence having at least 80% identity to PRO
 CC lacking its associated signal peptide, an extracellular domain of PRO
 CC with or without its associated signal peptide. Also included are vectors,
 CC transformed host cells, anti-PRO antibodies, the nucleic acids encoding
 CC PRO, PRO fusion proteins, inducing endothelial cell tube formation (by
 CC administering PRO281, PRO1560, PRO189, PRO4499, PRO6308, PRO6000,
 CC PRO10275, PRO21207, PRO20933 or PRO34274 polypeptide or its agonist) and
 CC an oligonucleotide probe derived from any one of the above nucleotide
 CC sequences. PRO6018 polypeptide is useful for stimulating the
 CC proliferation or differentiation of chondrocyte cells. PRO1313, PRO20080
 CC and PRO21383 polypeptides are useful for stimulating the proliferation of
 CC human microvascular endothelial cells. PRO6071, PRO4487 and PRO6006
 CC polypeptides are useful for inhibiting the proliferation of human
 CC microvascular endothelial cells. PRO polypeptides are useful for
 CC detecting the presence of tumour in a mammal, including tumours of lung,
 CC colon, breast, prostate, rectal, kidney and liver. PRO281, PRO1560,
 CC PRO189, PRO4499, PRO6308, PRO6000, PRO10275, PRO21207, PRO20933 and
 CC PRO34274 polypeptides are useful for inducing endothelial cell tube
 CC formation. PRO or the antibody are useful in the preparation of a
 CC medicament for treating a condition responsive to PRO polypeptide. The
 CC oligonucleotide probes are useful for isolating genomic and cDNA
 CC nucleotide sequences, for measuring or detecting the expression of an
 CC associated gene, and as antisense probes. PRO nucleic acid is useful as a
 CC hybridisation probe, in chromosome and gene mapping, in the generation of
 CC antisense RNA and DNA, and for the preparation PRO polypeptides. PRO and
 CC PRO nucleic acid are useful as therapeutic agents, e.g. vaccines. The
 CC present sequence encodes a PRO protein.

SQ Sequence 943 BP; 152 A; 306 C; 299 G; 186 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 114 Length: 943
 Score: 93.50 Matches: 36
 Percent Similarity: 35.25% Conservative: 7
 Best Local Similarity: 29.51% Mismatches: 31
 Query Match: 20.50% Indels: 48
 DB: 8 Gaps: 6

US-09-972-032-2 (1-79) x ADA43576 (1-943)

QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
 Db 138 TGTGCCAGCGCTGAGAGTGGCGCAGCGCAGCGCTGTGGAGTCTGCTGGC 197
 QY 22 LeuTrpThrGlyLeuGlyGluGlyGlnGluGlyGlyLeuGlyProGlu 37
 Db 198 TGTGGACAGGACCCCGGGAGGCGCA-----GCCCTGGGGCCAGAGCGGCCAGGTGG 251
 QY 38 -----GlyGlnAlaSerProThrPro 44
 Db 252 ACGCACATGACTGTAGCGCCTGGCTGGGGCTCCGAGGAGCGCGCTTCCAGAGTCC 311
 QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52
 Db 312 GAGGCACCGTCAAACTGCAGTTCGACATGATGGCGCTGCAACCTGGTGGCCACGCGCG 371
 QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66
 Db 372 CGTCTACCGCAGGCGCAGCTCACCTTCTCTGGGGCTGGTGGCGCTGCCCTGCTGTAC 431
 QY 66 GAspArgProGlnLeu-----GlyGlnLeuCysMe 76
 Db 432 CCGACGCGCGCTGTGGAGGAGGCCATGGCCGCTGCATCCAACTGGCGAGTTTGTCC 491

QY 76 tGly 77
 Db 492 TGGT 495

RESULT 12
 ADA01251
 ID ADA01251 standard; cDNA; 943 BP.
 AC ADA01251;
 DT 06-NOV-2003 (first entry)
 XX Human PRO polynucleotide #56.
 DE Human; gene; ss; PRO; secreted polypeptide; transmembrane polypeptide;
 KW tumour necrosis factor-alpha; TNF-alpha; blood; chondrocyte cell; tumour;
 KW cancer; lung; colon; breast; prostate; rectum; kidney; liver;
 KW microvascular endothelial cell; endothelial cell tube formation.
 XX Homo sapiens.
 OS
 PN US2003068782-A1.
 XX
 PD 10-APR-2003.
 XX
 PF 16-SEP-2002; 2002US-00245851.
 XX
 PR 27-APR-1999; 99US-0131271P.
 PR 29-OCT-1999; 99US-0162506P.
 PR 02-DEC-1999; 99WO-US028551.
 PR 29-AUG-2001; 2001WO-US027099.
 PR 18-JUL-2002; 2002US-00197942.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Baker KP, Eaton DL, Filvaroff E, Goddard A, Grimaldi JC;
 PI Gurney AL, Smith V, Stephan JP, Watanabe CK, Wood WI, Zhang Z;
 PI Fong S;
 XX
 DR WPI; 2003-625487/59.
 DR P-PSDB; ADA01252.
 XX
 XX Novel isolated PRO polypeptides e.g. PRO281 and PRO1560, useful in the
 PT preparation of a medicament for treating a condition responsive to PRO
 PT polypeptide, and as therapeutic agents e.g. vaccines.
 XX

Claim 2; Fig 111; 308pp; English.

PS The invention relates to isolated human PRO polypeptides (secreted and
 XX transmembrane polypeptides) and the polynucleotides encoding them. The
 CC invention also relates to an antibody which specifically binds to a PRO
 CC polypeptide, a method for stimulating the release of tumour necrosis
 CC factor-alpha (TNF-alpha) from human blood, a method for stimulating the
 CC proliferation or differentiation of chondrocyte cells and a method for
 CC detecting the presence of a tumour in a mammal (e.g. lung, colon, breast,
 CC prostate, rectal, kidney and liver tumours). The polynucleotides are
 CC useful in molecular biology, including uses as hybridisation probes, in
 CC chromosome and gene mapping, in generating antisense RNA and DNA and in
 CC gene therapy. The polynucleotides may also be used in preparing PRO
 CC polypeptides by recombinant techniques and in generating either
 CC transgenic animals or knock-out animals which are useful in the
 CC development and screening of therapeutically useful reagents. The PRO
 CC polypeptides or antibodies are used in preparing a medicament for
 CC treating a condition responsive to the polypeptides or antibodies, such
 CC as tumours, for stimulating and inhibiting proliferation of human
 CC microvascular endothelial cells and for inducing endothelial cell tube
 CC formation. This sequence represents a human PRO polynucleotide of the
 CC invention.

SQ Sequence 943 BP; 152 A; 306 C; 299 G; 186 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 114 Length: 943
 Score: 93.50 Matches: 36
 Percent Similarity: 35.25% Conservative: 7
 Best Local Similarity: 29.51% Mismatches: 31
 Query Match: 20.50% Indels: 48
 DB: 8 Gaps: 6

US-09-972-032-2 (1-79) x ADA01251 (1-943)

QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
 Db 138 TGTGCCAGCGCTGAGAGTGGCGCAGCGCAGCGCTGTGGAGTCTGCTGGC 197
 QY 22 LeuTrpThrGlyLeuGlyGluGlyGlnGluGlyGlyLeuGlyProGlu----- 37
 Db 198 TGTGGACAGGACCCCGGGAGGCGCA-----GCCCTGGGGCCAGAGCGGCCAGGTGG 251
 QY 38 -----GlyGlnAlaSerProThrPro 44
 Db 252 ACGCACATGACTGTAGCGCCTGGGGCTCCGAGGAGCGCGCTTCCAGAGTCC 311
 QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52
 Db 312 GAGGCACCGTCAAACTGCAGTTCGACATGATGGCGCTGCAACCTGGTGGCCACGCGCG 371
 QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66
 Db 372 CGTCTACCGCAGGCGCAGCTCACCTTCTCTGGGGCTGGTGGCGCTGCCCTGCTGTAC 431
 QY 66 GAspArgProGlnLeu-----GlyGlnLeuCysMe 76
 Db 432 CCGACGCGCGCTGTGGAGGAGGCCATGGCCGCTGCATCCAACTGGCGAGTTTGTCC 491

QY 76 tGly 77
 Db 492 TGGT 495

RESULT 13
 ADA01135
 ID ADA01135 standard; cDNA; 943 BP.
 XX
 AC ADA01135;
 XX
 DT 06-NOV-2003 (first entry)
 XX
 DE Human cDNA encoding secreted/transmembrane polypeptide PRO21434.
 XX

KW ss; gene; human; PRO; secreted protein; transmembrane protein;
 KW endothelial cell tube formation; chondrocyte cell differentiation;
 KW microvascular endothelial cell; tumour; lung tumour; colon tumour;
 KW breast tumour; prostate tumour; rectal tumour; kidney tumour;
 KW liver tumour; cytostatic; vaccine.
 XX Homo sapiens.
 XX OS
 XX PN US2003068780-A1.
 XX PD 10-APR-2003.
 XX PF 16-SEP-2002; 2002US-00245143.
 XX 02-AUG-2000; 2000US-0222695P.
 XX PR 20-JUN-2001; 2001WO-US019692.
 XX PR 28-AUG-2001; 2001WO-US027099.
 XX PR 18-JUL-2002; 2002US-00197942.
 XX (GETH) GENENTECH INC.
 XX PA Baker KP, Eaton DL, Filvaroff E, Goddard A, Grimaldi JC;
 PI Gurney AL, Smith V, Stephan JP, Watanabe CK, Wood WI, Zhang Z;
 PI Pong S;
 XX WPI; 2003-625485/59.
 DR P-PSDB; ADA01136.
 XX Novel isolated PRO polypeptides e.g. PRO281 and PRO1560, useful in the
 PT preparation of a medicament for treating a condition responsive to PRO
 PT polypeptide, and as therapeutic agents e.g. vaccines.
 XX Claim 2; Fig 111; 307pp; English.
 XX The invention relates to an isolated secreted/transmembrane (PRO)
 CC polypeptide, having at least 80% sequence identity to a sequence selected
 CC from any one of the 57 amino acid sequences given in specification, or to
 CC a sequence encoded by a nucleic acid molecule selected from any one of
 CC the nucleic acids deposited under any of the ATCC accession numbers given
 CC in specification, or a sequence having at least 80% identity to PRO
 CC lacking its associated signal peptide, an extracellular domain of PRO
 CC with or without its associated signal peptide. Also included are vectors,
 CC transformed host cells, anti-PRO antibodies, the nucleic acids encoding
 CC PRO, PRO fusion proteins, inducing endothelial cell tube formation (by
 CC administering PRO281, PRO1560, PRO189, PRO4499, PRO6308, PRO6000,
 CC PRO10275, PRO21207, PRO20933 or PRO34274 polypeptide or its agonist) and
 CC an oligonucleotide probe derived from any one of the above nucleotide
 CC sequences. PRO6018 polypeptide is useful for stimulating the
 CC proliferation or differentiation of chondrocyte cells. PRO1313, PRO20080
 CC and PRO21383 polypeptides are useful for stimulating the proliferation of
 CC human microvascular endothelial cells. PRO6071, PRO4487 and PRO6006
 CC polypeptides are useful for inhibiting the proliferation of human
 CC microvascular endothelial cells. PRO polypeptides are useful for
 CC detecting the presence of tumour in a mammal, including tumours of lung,
 CC colon, breast, prostate, rectal, kidney and liver. PRO281, PRO1560,
 CC PRO189, PRO4499, PRO6308, PRO6000, PRO10275, PRO21207, PRO20933 and
 CC PRO34274 polypeptides are useful for inducing endothelial cell tube
 CC formation. PRO or the antibody are useful in the preparation of a
 CC medicament for treating a condition responsive to PRO polypeptide. The
 CC oligonucleotide probes are useful for isolating genomic and cDNA
 CC nucleotide sequences, for measuring or detecting the expression of an
 CC associated gene, and as antisense probes. PRO nucleic acid is useful as a
 CC hybridisation probe, in chromosome and gene mapping, in the generation of
 CC antisense RNA and DNA, and for the preparation PRO polypeptides. PRO and
 CC PRO nucleic acid are useful as therapeutic agents, e.g. vaccines. The
 CC present sequence encodes a PRO protein.
 XX SQ Sequence 943 BP; 152 A; 306 C; 299 G; 186 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 114 Length: 943
 Score: 93.50 Matches: 36
 Percent Similarity: 35.25% Conservative: 7

Best Local Similarity: 29.51% Mismatches: 31
 Query Match: 20.50% Indels: 48
 DB: 8 Gaps: 6
 US-09-972-032-2 (1-79) x ADA01135 (1-943)
 QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
 DB 138 TGTGCCAGACGCTGGAGATGGCGCAGGCGAGCGGCTGTGGAGGTCTCTGTGTC 197
 QY 22 LeuTrpThrGlyLeuGlyGlnGluGlyGlyVileGlyProGlu----- 37
 DB 198 TGTGGACAGGACCCGGGAGGGCCGA-----GCCCTGGGGCAGAGCGGCGAGTGG 251
 QY 38 -----GlyGlnAlaSerProThrPro 44
 DB 252 ACGCACATGACTGTGAGCGCGCTGGGCTGGGGCTCCGAGGCGAGCGGCTCCAGGAGTCCC 311
 QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52
 DB 312 GAGCACCGTCAACATGACATGATGCGCGCTCAACCTGGTGGCCACGCGCG 371
 QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66
 DB 372 CGCTCACCGCAGCGACGTCACCTTCCTCTCGGGGCTGGTGGCGCTGCCCTGCTGTCAC 431
 QY 66 gAspArgProGlnLeu-----GlyGluLeuCysMe 76
 DB 432 CCGACGCGCGCGTGTGGGAGGAGGCCATGGCCGCTGCATTCACACTGGCGAGTTTGTC 491
 QY 76 tGly 77
 DB 492 TGGT 495
 RESULT 14
 ADA43692
 ID ADA43692 standard; cDNA; 943 BP.
 XX AC ADA43692;
 XX DT 20-NOV-2003 (first entry)
 XX DE Human cDNA encoding secreted/transmembrane polypeptide PRO21434.
 XX KW ss; Gene; human; PRO; secreted protein; transmembrane protein;
 KW endothelial cell tube formation; chondrocyte cell differentiation;
 KW microvascular endothelial cell; tumour; lung tumour; colon tumour;
 KW breast tumour; prostate tumour; rectal tumour; kidney tumour;
 KW liver tumour; cytostatic; vaccine.
 XX OS Homo sapiens.
 XX PN US2003073190-A1.
 XX PD 17-APR-2003.
 XX PF 09-SEP-2002; 2002US-00238283.
 XX PR 01-JUL-1998; 98US-0091358P.
 XX PR 01-JUN-1999; 99WO-US012252.
 XX PR 20-JUL-1999; 99US-0144758P.
 XX PR 28-JUL-1999; 99US-0146222P.
 XX PR 25-AUG-1999; 99US-00380137.
 XX PR 30-MAR-2000; 2000WO-US008439.
 XX PR 02-JUN-2000; 2000WO-US015264.
 XX PR 29-AUG-2001; 2001WO-US027099.
 XX PR 18-JUL-2002; 2002US-00197942.
 XX (GETH) GENENTECH INC.
 XX PA Baker KP, Eaton DL, Filvaroff E, Goddard A, Grimaldi JC;
 PI Gurney AL, Smith V, Stephan JP, Watanabe CK, Wood WI, Zhang Z;
 PI Pong S;

XX WPI; 2003-585304/55.
 DR P-PSDB; ADA43693.
 XX
 PT New isolated nucleic acid encoding a PRO polypeptide, e.g. PRO20080 or
 PT PRO21383, useful in molecular biology, chromosome and gene mapping, in
 PT generating antisense RNA and DNA, and in gene therapy.
 XX
 PS Claim 2; Fig 111; 352pp; English.
 XX
 CC The invention relates to an isolated secreted/transmembrane (PRO)
 CC polypeptide, having at least 80% sequence identity to a sequence selected
 CC from any one of the 57 amino acid sequences given in specification, or to
 CC a sequence encoded by a nucleic acid molecule selected from any one of
 CC the nucleic acids deposited under any of the ATCC accession numbers given
 CC in specification, or a sequence having at least 80% identity to PRO
 CC lacking its associated signal peptide, an extracellular domain of PRO
 CC with or without its associated signal peptide. Also included are vectors,
 CC transformed host cells, anti-PRO antibodies, the nucleic acids encoding
 CC PRO, PRO fusion proteins, inducing endothelial cell tube formation (by
 CC administering PRO281, PRO1560, PRO189, PRO4499, PRO6308, PRO6000;
 CC PRO10275, PRO21207, PRO20933 or PRO34274 polypeptide or its agonist) and
 CC an oligonucleotide probe derived from any one of the above nucleotide
 CC sequences. PRO6018 polypeptide is useful for stimulating the
 CC proliferation or differentiation of chondrocyte cells. PRO1313, PRO20080
 CC and PRO21383 polypeptides are useful for stimulating the proliferation of
 CC human microvascular endothelial cells. PRO6071, PRO4487 and PRO6006
 CC polypeptides are useful for inhibiting the proliferation of human
 CC microvascular endothelial cells. PRO polypeptides are useful for
 CC detecting the presence of tumour in a mammal, including tumours of lung,
 CC colon, breast, prostate, rectal, kidney and liver. PRO281, PRO1560,
 CC PRO189, PRO4499, PRO6308, PRO6000, PRO10275, PRO21207, PRO20933 and
 CC PRO34274 polypeptides are useful for inducing endothelial cell tube
 CC formation. PRO or the antibody are useful in the preparation of a
 CC medicament for treating a condition responsive to PRO polypeptide. The
 CC oligonucleotide probes are useful for isolating genomic and cDNA
 CC associated genes, for measuring or detecting the expression of an
 CC associated gene, and as antisense probes. PRO nucleic acid is useful as a
 CC hybridisation probe, in chromosome and gene mapping, in the generation of
 CC antisense RNA and DNA, and for the preparation PRO polypeptides. PRO and
 CC PRO nucleic acid are useful as therapeutic agents, e.g. vaccines. The
 CC present sequence encodes a PRO protein.
 XX
 SQ Sequence 943 BP; 152 A; 306 C; 299 G; 186 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 114 Length: 943
 Score: 93.50 Matches: 36
 Percent Similarity: 35.25% Conservative: 7
 Best Local Similarity: 29.51% Mismatches: 31
 Query Match: 20.50% Indels: 48
 DB: Gaps: 6

US-09-972-032-2 (1-79) x ADA43692 (1-943)

QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
 DB 138 TGTGCCAGACGCTGGAGAGATGGCGCAGCGCGAGCTGGGCTGTGGAGTCTGCTGTC 197
 QY 22 LeuTrpThrGlyLeuGlyGluGlyGlnGluGlyGlyLeuGlyProGlu----- 37
 DB 198 TGTGTGACAGACCCGGGGAGGCCCA-----GCCCTGGGGCCAGAGCGGCCAGGTGG 251
 QY 38 -----GlyGlnAlaSerProThrPro 44
 DB 252 ACGCACATGACTGTAGGCGCTGGGCTGGGGCTCCGAGGACGCGCGCTTCCAGGAGTCCC 311
 QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52
 DB 312 GAGGACCGTCAAACTGCAGTTCGACATGATGCGGCCCTGCAACCTGGGGCCAGCGCG 371
 QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66

Db 372 CGCTCACCGCAGGCGAGCTCACCTTCCTCTGGGGCTGGTGGGCGCTGCCCTGTGTGTAC 431
 QY 66 gAsPArgProGlnLeu-----GlyGluLeuCysMe 76
 DB 432 CCGAGCGCCCGTGTGGGAGGAGGCCATGGCGCGTGCATTCCAACTGCGGAGTTTGTCC 491
 QY 76 tGly 77
 DB 492 TGGT 495

RESULT 15
 ADA06954
 ID ADA06954 standard; cDNA; 943 BP.
 XX ADA06954;
 AC ADA06954;
 XX
 DT 06-NOV-2003 (first entry)
 XX
 DE Human PRO polynucleotide #56.
 XX
 KW Human; PRO; gene; ss; secreted polypeptide; transmembrane polypeptide;
 KW tumour necrosis factor-alpha; TNF-alpha; blood; chondrocyte cell; tumour;
 KW adrenal; lung; colon; breast; prostate; rectum; cervix; liver;
 KW microvascular endothelial cell; endothelial cell tube formation;
 KW bone disorder; cartilage disorder; sports injury; proteoglycan;
 KW cartilage; sports-related joint problem; articular cartilage defect;
 KW osteoarthritis; rheumatoid arthritis; haemoglobin-associated disorder;
 KW thalassemia; immune system cell infiltration; cancer; vulnery;
 KW antianemic; osteopathic; antirheumatic; antiarthritic.
 XX
 OS Homo sapiens.
 XX
 PN US2003068781-A1.
 PN
 PD 10-APR-2003.
 XX
 PF 16-SEP-2002; 2002US-00245771.
 XX
 PR 03-AUG-1999; 99US-0146843P.
 PR 15-MAY-2000; 2000WO-US013358.
 PR 29-AUG-2001; 2001WO-US027099.
 PR 18-JUL-2002; 2002US-00197942.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Baker KP, Eaton DL, Filvaroff E, Goddard A, Grimaldi JC;
 PI Gurney AL, Smith V, Stephan JP, Watanabe CK, Wood WL, Zhang Z;
 PI Fong S;
 XX
 DR WPI; 2003-625486/59.
 DR P-PSDB; ADA06955.
 XX
 PT Novel secreted and transmembrane polypeptides, PRO polypeptides useful
 PT for stimulating proliferation or differentiation of chondrocyte cells and
 PT inducing endothelial cell tube formation.
 XX
 PS Claim 2; Fig 111; 307pp; English.
 XX
 CC The invention relates to isolated human PRO polypeptides (secreted and
 CC transmembrane polypeptides) and the polynucleotides encoding them. The
 CC invention also relates to an antibody which specifically binds to a PRO
 CC polypeptide, a method for stimulating the release of tumour necrosis
 CC factor-alpha (TNF-alpha) from human blood, a method for stimulating the
 CC proliferation or differentiation of chondrocyte cells and a method for
 CC detecting the presence of a tumour in a mammal (e.g. adrenal, lung,
 CC colon, breast, prostate, rectal, cervical and liver tumours). The
 CC polynucleotides are useful in molecular biology, including uses as
 CC hybridisation probes, in chromosome and gene mapping, in generating
 CC antisense RNA and DNA and in gene therapy. The polynucleotides may also
 CC be used in preparing PRO polypeptides by recombinant techniques and in
 CC generating either transgenic animals or knock-out animals which are
 CC useful in the development and screening of therapeutically useful
 CC reagents. The PRO polypeptides or antibodies are used in preparing a

CC medicament for treating a condition responsive to the polypeptides or
 CC antibodies, such as tumours, for stimulating and inhibiting proliferation
 CC of human microvascular endothelial cells, for inducing endothelial cell
 CC tube formation and for treating various bone and/or cartilage disorders
 CC such as sports injuries and arthritis. PRO polypeptides which stimulate
 CC the release of proteoglycans from cartilage are useful for treating
 CC osteoarthritis and rheumatoid arthritis. PRO polypeptides are also useful
 CC for treating various mammalian haemoglobin-associated disorders such as
 CC various thalassaemias and conditions which may benefit from enhanced
 CC local immune system cell infiltration. This sequence represents a human
 CC PRO polynucleotide of the invention.

XX
 SQ Sequence 943 BP; 152 A; 306 C; 299 G; 186 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 114 Length: 943
 Score: 93.50 Matches: 36
 Percent Similarity: 35.25% Conservatives: 7
 Best Local Similarity: 29.51% Mismatches: 31
 Query Match: 20.50% Indels: 48
 DB: 8 Gaps: 6

US-09-972-032-2 (1-79) x ADA06954 (1-943)

Qy 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
 Db 138 TGTGCCAGACGCTGGAGGATGGCGCAGCGCGAGCGTGGGGCTGTGGAGGTCTCTGCTGC 197
 Qy 22 LeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlyLeuGlyProGlu----- 37
 Db 198 TGGTGGACAGACCCGGGGAGGCCGA-----GCCCTGGGGCCAGCGCCGCCAGGTGG 251
 Qy 38 -----GlyGlnAlaSerProThrPro 44
 Db 252 ACGCACATGACTGTGAGGCGCTGGGCTGGGGCTCCGAGGCGCGCTTCCAGGAGTCCC 311
 Qy 45 Asp-----CysAlaSerArg-----TrpProArg--- 52
 Db 312 GAGGACCCGTCAACTGCAGTTCGACATGCGCGCCTGCAACCTGGTGGCCACGCGCG 371
 Qy 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValar 66
 Db 372 CGCTCACCGCAGCGCAGCTACCTTCTCTGGGGCTGGTGGGCTGCCCCCTGCTGTAC 431
 Qy 66 gAspArgProGlnLeu-----GlyGlnLeuCysMe 76
 Db 432 CCGACGCCCGCTGTGGAGGAGGCCATGCGCGCTCATTCCAACTGGCGAGTTTGTGCC 491
 Qy 76 tGly 77
 Db 492 TGGT 495

Search completed: July 10, 2004, 16:58:44
 Job time : 452 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 13, 2004, 08:38:15 ; Search time 2883 Seconds
(without alignments)
818.284 Million cell updates/sec

Title: US-09-972-032-2
Perfect score: 79
Sequence: 1 MCGRPVRSVSGCFADAHWT.....SAGLTVRDRLGELCMGRG 79

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Word size: 1

Total number of hits satisfying chosen parameters: 55023952

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09972032/runat_06072004_121534_8884/app query.fasta_1.263
-DB=EST -QPMI=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=plo
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000.
-USPR=US09972032 @CNG 1.1 3437 @runat_06072004_121534_8884 -NCPU=6 -ICPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -USPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mas:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_pbg:*
27: em_gss_vrl:*
28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	79	100.0	1467	11	BC039500	BC039500 Homo sapi
2	45	57.0	691	13	BU620427	BU620427 UI-H-FL1-
3	44	55.7	1201	13	EX406138	EX406138 BX406138
4	34	43.0	669	13	BM679577	BM679577 UI-E-E00-
5	34	43.0	694	13	BU684594	BU684594 UI-CF-EN1
6	34	43.0	760	12	BM980506	BM980506 UI-CF-EN1
7	33	41.8	632	13	BU617331	BU617331 UI-H-DF0-
8	33	41.8	633	14	CA313142	CA313142 UI-CF-FN0
9	32	40.5	623	14	CD742581	CD742581 UI-H-FT2-
10	23	29.1	628	13	EQ447041	EQ447041 UI-H-EU1-
11	22	27.8	1084	12	BM806108	BM806108 AGENCOURT
12	17	21.5	282	10	BF364571	BF364571 PMI-NM108
13	9	11.4	163	14	R86559	R86559 RABEST126T
14	9	11.4	181	9	AA656691	AA656691 VP9SC01.R
15	9	11.4	424	9	AA593543	AA593543 mn28C03.s
16	9	11.4	526	9	A1022772	A1022772 ow99b10.x
17	9	11.4	554	14	CB471798	CB471798 sn45_A09.
18	9	11.4	627	13	BU405769	BU405769 604138917
19	9	11.4	691	13	BU113858	BU113858 603131524
20	9	11.4	750	13	BU403370	BU403370 604138868
21	9	11.4	793	13	BU357204	BU357204 603475234
22	9	11.4	838	13	BU354677	BU354677 603475426
23	9	11.4	1047	10	BE621897	BE621897 601494186
24	9	11.4	1076	10	BF168132	BF168132 601776390
25	8	10.1	134	28	AZ099785	AZ099785 RPCI-23-4
26	8	10.1	172	10	AW384572	AW384572 IL2-HT039
27	8	10.1	172	14	W30123	W30123 mc27B09.r1
28	8	10.1	176	13	EQ761225	EQ761225 EBR004.SQ
29	8	10.1	187	28	BH23431	BH23431 1006113A0
30	8	10.1	197	14	W14247	W14247 mb27f07.r1
31	8	10.1	201	29	CE017568	CE017568 tigr-gss-
32	8	10.1	215	10	BB293476	BB293476 BB293476
33	8	10.1	218	29	CE660402	CE660402 tigr-gss-
34	8	10.1	220	9	A1005741	A1005741 ua74h11.r
35	8	10.1	248	28	AQ113700	AQ113700 CIR-HSP-2
36	8	10.1	252	28	AZ577620	AZ577620 12a02 Sho
37	8	10.1	272	10	BF721038	BF721038 mab62a09.
38	8	10.1	272	13	BQ807616	BQ807616 NISC_kk06
39	8	10.1	279	10	BF651626	BF651626 274459 MA
40	8	10.1	279	10	BF012245	BF012245 u640a01.Y
41	8	10.1	280	14	F27454	F27454 HSPD15401.H
42	8	10.1	283	10	BE717721	BE717721 RC4-HT078
43	8	10.1	283	10	BE837703	BE837703 RC2-FN009
44	8	10.1	284	14	CD810225	CD810225 957 Lingu
45	8	10.1	286	10	BE837692	BE837692 RC2-FN009

ALIGNMENTS

RESULT 1
BC039500
LOCUS BC039500 1467 bp mRNA HTC 04-MAR-2003
DEFINITION Homo sapiens, clone IMAGE:5555626, mRNA.
ACCESSION BC039500
VERSION BC039500.1 GI:25058499
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1467)
AUTHORS Strausberg,R.


```

VERSION      BX406138.1  GI:30648317
KEYWORDS     Homo sapiens (human)
SOURCE       Homo sapiens
ORGANISM     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1 (bases 1 to 1201)
AUTHORS      Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE        Full-length cDNA libraries and normalization
JOURNAL      Unpublished (2001)
COMMENT      Contact: Genoscope
              Genoscope - Centre National de Sequencage
              BP 191 91006 EVRY cedex - France
              Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
              Library was constructed by Life Technologies, a division of
              Invitrogen. This sequence belongs to sequence cluster 8395.f For
              more information about this cluster, see
              http://www.genoscope.cns.fr/
              cgi-bin/cluster.cgi?seq=CSOAM010CB10NP1&cluster=8395.f. Contact :
              Peng Liang Email : fliang@lifetech.com URL :
              http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
              Faraday Avenue Genoscope sequence ID : CSOAM010CB10NP1.
FEATURES     Location/Qualifiers
              1..1201
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="CSOAM010YD19"
                /tissue_type="FETAL LIVER"
                /dev_stage="fetal"
                /clone_lib="Homo sapiens FETAL LIVER"
                /notes="Organ: liver; Vector: pCMVSPORT 6; 1st strand cDNA
              was primed with a NotI-oligo(dT) primer. Five prime end
              enriched, double-strand cDNA was digested with Not I and
              cloned into the Not I and EcoRV sites of the pCMVSPORT 6
              vector. Library was not normalized."
ORIGIN
Alignment Scores:
Pred. No.:      5,05e-29      Length:      1201
Score:          44.00      Matches:      65
Percent Similarity: 98.48%      Conservative: 0
Best Local Similarity: 98.48%      Mismatches: 1
Query Match:    55.70%      Indels:      1
DB:             13          Gaps:         0

US-09-972-032-2 (1-79) x BX406138 (1-1201)
Qy 14 PhenAlaSerArgTrpThrGlyLeuTrpThrGlyLeuGlyGlnGlyGly 33
Db 674 TTCGCTGATGCATTCGACCGGGCTCTGGACTGGGCTAGGGAAGGCGAGGGCGGA 615
Qy 34 IleGlyProGluGlyGlnAlaSerProThrProAspCysAlaSerArgTrpProArgSer 53
Db 614 ATTGG-CCCGAGGGCCAGGCTCGCCACCCCGACTGGCCCTCCCGTGGCCCGCAGC 556
Qy 54 AlaSerArgTrpProTrpSerAlaGlyLeuThrValArgAspArgProGlnLeuGlyGlu 73
Db 555 GCCCTCCCGTGGCCCTGGAGTGCAGGTCCTTACCGTCCGAGATCGTCGCAACTGGGCGAG 496
Qy 74 LeuCysMetGlyArgGly 79
Db 495 CTGTGTCATGGGGCGTGGC 478

RESULT 4
LOCUS      BM679577/c
DEFINITION UI-E-E00-aia-1-05-0-UI.s1 UI-E-E00 Homo sapiens cDNA clone
            669 bp mRNA linear EST 27-FEB-2002
ACCESSION  BM679577
VERSION    BM679577.1 GI:18989473
KEYWORDS   EST.
SOURCE     Homo sapiens (human)

ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 669)
AUTHORS    Ronaldo, M.F., Lennon, G. and Soares, M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
              discovery
JOURNAL    Genome Res. 6 (9), 791-806 (1996)
MEDLINE    97044477
PUBMED     889548
COMMENT     Contact: Soares, MB
              Coordinated Laboratory for Computational Genomics
              University of Iowa
              375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
              Tel: 319 335 8250
              Fax: 319 335 9585
              Email: bento-soares@uiowa.edu
              Tissue Procurement: Dr. Gregg Hageman
              cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
              cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
              DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
              Clone Distribution: Researchers may obtain clones from Research
              Genetics (www.resgen.com).
              Seq primer: M13 Forward
              POLYA=Yes.
              Location/Qualifiers
              1..669
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="UI-E-E00-aia-1-05-0-UI"
                /tissue_type="fetal eye"
                /dev_stage="fetal"
                /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                /clone_lib="UI-E-E00"
                /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
              modified polylinker; Site 1: EcoR I; Site 2: Not I;
              UI-E-E00 is a cDNA library containing the following
              tissue(s): fetal eye. The library was constructed
              according to Ronaldo, Lennon and Soares, Genome Research,
              6:791-806, 1996. First strand cDNA synthesis was primed
              with an oligo-dT primer containing a Not I site. Double
              stranded cDNA was ligated to an EcoR I adaptor, digested
              with Not I, and cloned directionally into pT7T3-Pac
              vector. The oligonucleotide used to prime the synthesis of
              first-strand cDNA contains a library tag sequence that is
              located between the Not I site and the (dT)18 tail. The
              sequence tag for this library is CCGGTATACC. This library
              was created for the program, Gene Discovery in the Visual
              System, supported by National Eye Institute (NEI).
              TAG_TISSUE=human fetal eye
              TAG_LIB=UI-E-E00
              TAG_SEQ=CCGGTATACC"
ORIGIN
Alignment Scores:
Pred. No.:      2.33e-20      Length:      569
Score:          34.00      Matches:      34
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    43.04%      Indels:      0
DB:             12          Gaps:         0

US-09-972-032-2 (1-79) x BM679577 (1-669)
Qy 46 CysAlaSerArgTrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThrVal 65
Db 642 TCGCGCTCCCGTGGCCCGCGCAGCGCTCCCGTGGCCCTCGAGTGCAGGCTTACCGTCC 583
Qy 66 ArgAspArgProGlnLeuGlyGlnLeuCysMetGlyArgGly 79
Db 582 CGAGATCGTCCGCAACTGGCGAGCTGTGTCATGGGGCGGTGGC 541

```

```

RESULT 5
BU684594/c
LOCUS
DEFINITION
  UI-CF-EN1-act-a-22-0-UI.s1 694 bp mRNA linear EST 07-OCT-2002
  UI-CF-EN1-act-a-22-0-UI 3', mRNA sequence.
ACCESSION
  BU684594
VERSION
  BU684594.1 GI:23537704
SOURCE
  EST.
ORGANISM
  Homo sapiens (human)
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  Bonaldo,M.F., Lennon,G. and Soares,M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
  Genome Res. 6 (9), 791-806 (1996)
  97044477
  8889548
  McCray Lab
  University of Iowa
  2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
  Tel: 319 356 4866
  Fax: 319 356 7171
  Email: paul-mccray@uiowa.edu
  Tissue Procurement: Dr. M. J. Welsh, University of Iowa
  CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
  CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Researchers may obtain clones from Research
  Genetics (www.resgen.com) or from Open Biosystems
  (www.openbiosystems.com).
  Seq primer: M13 FORWARD
  POLYA=Yes.

FEATURES
  source
  1..694
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="UI-CF-EN1-act-a-22-0-UI"
    /tissue_type="Primary Lung Cystic Fibrosis Epithelial
    Cells"
    /dev_stage="Adult"
    /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
    /clone_lib="UI-CF-EN1"
    /notes="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
    modified polylinker; Site 1: EcoR I; Site 2: Not I;
    UI-CF-EN1 is a normalized cDNA library containing the
    following tissue(s): Primary Lung Cystic Fibrosis
    Epithelial Cells. The library was constructed according to
    Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
    1996. First strand cDNA synthesis was primed with an
    oligo-dT primer containing a Not I site. Double stranded
    cDNA was ligated to an EcoR I adaptor, digested with Not
    I, and cloned directionally into pT7T3-Pac vector. The
    oligonucleotide used to prime the synthesis of
    first-strand cDNA contains a library tag sequence that is
    located between the Not I site and the (dT)18 tail. The
    sequence tag for this library is CTCCTCAGGT.
    TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS
    6hr to LPS 24h
    TAG LIB=UI-CF-EN1
    TAG_SEQ=CTGCTCAGGT"

ORIGIN
Alignment Scores:
Pred. No.: 2,42e-20 Length: 694
Score: 34.00 Matches: 34
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 43.04% Indels: 0
DB: 13 Gaps: 0

```

```

US-09-972-032-2 (1-79) x BU684594 (1-694)
QY 46 CysAlaSerArgTrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThrVal 65
DB 647 TCGCGCTCCGGTGGCCCGCAGCGCCCTCCGGTGGCCCTCCGGTGGCAGGTCTTACCGTC 588
QY 66 ArgAspArgProGlnLeuGlyGlyLeuGlyMetGlyArgGly 79
DB 587 CGAGATGCTCCGCACTGGCGAGCTGTGCATGGGGCGTGGC 546
RESULT 6
LOCUS
DEFINITION
  UI-CF-EN1-add-a-08-0-UI.s1 760 bp mRNA linear EST 21-FEB-2003
  UI-CF-EN1-add-a-08-0-UI 3', mRNA sequence.
ACCESSION
  BM980506
VERSION
  BM980506.1 GI:19602038
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  Bonaldo,M.F., Lennon,G. and Soares,M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
  Genome Res. 6 (9), 791-806 (1996)
  97044477
  8889548
  McCray Lab
  University of Iowa
  2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
  Tel: 319 356 4866
  Fax: 319 356 7171
  Email: paul-mccray@uiowa.edu
  Tissue Procurement: Dr. M. J. Welsh, University of Iowa
  CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
  CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Researchers may obtain clones from Research
  Genetics (www.resgen.com) or from Open Biosystems
  (www.openbiosystems.com).
  Seq primer: M13 FORWARD
  POLYA=Yes.

```

```

FEATURES
  source
  1..760
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="UI-CF-EN1-add-a-08-0-UI"
    /tissue_type="Primary Lung Cystic Fibrosis Epithelial
    Cells"
    /dev_stage="Adult"
    /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
    /clone_lib="UI-CF-EN1"
    /notes="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
    modified polylinker; Site 1: EcoR I; Site 2: Not I;
    UI-CF-EN1 is a normalized cDNA library containing the
    following tissue(s): Primary Lung Cystic Fibrosis
    Epithelial Cells. The library was constructed according to
    Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
    1996. First strand cDNA synthesis was primed with an
    oligo-dT primer containing a Not I site. Double stranded
    cDNA was ligated to an EcoR I adaptor, digested with Not
    I, and cloned directionally into pT7T3-Pac vector. The
    oligonucleotide used to prime the synthesis of
    first-strand cDNA contains a library tag sequence that is
    located between the Not I site and the (dT)18 tail. The
    sequence tag for this library is CTCCTCAGGT.
    TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS
    6hr to LPS 24h

```

TAG_LIB=UI-CF-ENI
TAG_SEQ=CTGCTCAGGT

ORIGIN

Alignment Scores:
Pred. No.: 2.56e-20 Length: 760
Score: 34.00 Matches: 65
Percent Similarity: 97.01% Conservative: 0
Best Local Similarity: 97.01% Mismatches: 1
Query Match: 43.04% Indels: 2
DB: 12 Gaps: 0

US-09-972-032-2 (1-79) x BM980506 (1-760)

Qy 14 PheAlaAspAlaHisTrpThrGlyLeuTrpThrGlyLeuGlyGlnGlnGlyGly 33
Db 732 TTCGCTGATGACATTTGGACCGGCTCTGACTGGGCTAGGGAGGAGGAGGCGGA 673
Qy 34 IleGlyProGlnGlyGlnAlaSerProThr-ProAspCysAlaSerArgTrpProArgSe 53
Db 672 ATTGGGCGGAGGCGGAGGCTCGCCGACCCCGAN-TGGCCCTCCGGTGGCCCGCAG 614
Qy 53 rAlaSerArgTrpProTrpSerAlaGlyLeuThrValArgAspArgProGlnLeuGlyG1 73
Db 613 CGCTCCCGGTGGCCCTGGAGTGCAGTCTTACCGTCGAGATCGTCCGCAACTGGGCGA 554
Qy 73 uLeuCysMetGlyArgGly 79
Db 553 GCTGTGCATGGGCGGTGGC 535

RESULT 7

BU617331/c
LOCUS BU617331 632 bp mRNA linear EST 23-SEP-2002
DEFINITION UI-H-DF0-bep-n-09-0-UI.s1 NCI_CGAP_DFO Homo sapiens cDNA clone
UI-H-DF0-bep-n-09-0-UI 3', mRNA sequence.
ACCESSION BU617331
VERSION BU617331.1 GI:23283539
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 632)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES

source
Location/Qualifiers
1..632
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-DF0-bep-n-09-0-UI"
/tissue_type="Subchondral Bone"
/dev_stage="Adult"
/lab_hosts="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP_DFO"
/note="Organ: Bone; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site: 1; Ecor I; Site: 2; Not 1;
NCI CGAP_DFO is a cDNA library containing the following
tissue(s): Subchondral Bone. The library was constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed

ORIGIN

Alignment Scores:
Pred. No.: 1.71e-19 Length: 632
Score: 33.00 Matches: 33
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 41.77% Indels: 0
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x BU617331 (1-632)

Qy 47 AlaSerArgTrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThrValArg 66
Db 630 GCCTCCCGGTGGCCCGCAGCGCTCCCGGTGGCCCTGGAGTGCAGGTCTTACCGTCCGA 571
Qy 67 AspArgProGlnLeuGlyGlnLeuCysMetGlyArgGly 79
Db 570 GATCGTCCGCAACTGGCGAGCTGTGATGGGCGTGGC 532

RESULT 8

CA313142/c
LOCUS CA313142 633 bp mRNA linear EST 04-NOV-2002
DEFINITION UI-CF-FNO-aex-n-23-0-UI.s1 UI-CF-FNO Homo sapiens cDNA clone
UI-CF-FNO-aex-n-23-0-UI 3', mRNA sequence.
ACCESSION CA313142
VERSION CA313142.1 GI:24531240
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 633)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES

Location/Qualifiers
1..633
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-FNO-aex-n-23-0-UI"
/tissue_type="Human Lung Epithelial cells"
/lab_hosts="DH10B (Life Technologies) (T1 phage resistant)"

with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is GTTAAGCGTC.

TAG_LIB=UI-H-DF0

TAG_SEQ=GTTAAGCGTC

/clone_lib="UI-CF-FNO"
 /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 UI-CF-FNO is a subcloned cDNA library derived from two
 normalized Human lung epithelial cell libraries (EN1 and
 DUL) The library was subcloned according to according to
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
 1996. For additional information, contact:
 bento-soares@uiowa.edu
 TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS
 5hr to LPS 24h
 TAG LIB=UI-CF-FNO
 TAG_SEQ=CTGCTCAGGT"

ORIGIN

Alignment Scores:
 Pred. No.: 1.72e-19 Length: 633
 Score: 33.00 Matches: 33
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 41.77% Indels: 0
 DB: 14 Gaps: 0

US-09-972-032-2 (1-79) x CA313142 (1-633)

QY 47 AlasArGrTpProArGrSerAlasArGrTpProTrpSerAlaglyLeuThrValArg 66
 |||||
 Db 633 GCCTCCCGTGGCCCGCAGCGCTCCCGTGGCCCTGGAGTGCAGGTCTTACCGTCCGA 574
 |||||
 QY 67 AspArGrProGlnLeuGlyGluLeuCysMetGlyArgGly 79
 |||||
 Db 573 GATCGTCCGCACTGGCGAGCTGTGCATGGCGCGTGGC 535
 |||||

RESULT 9

CD742581/c
 LOCUS
 DEFINITION UI-H-FT2-bj1-i-10-0-UI.s1 NCI CGAP FT2 Homo sapiens cDNA clone
 UI-H-FT2-bj1-i-10-0-UI 3', mRNA sequence.
 CD742581
 VERSION
 KEYWORDS EST.
 SOURCE CD742581.1 GI:32293431
 Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 623)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Dr. Gary W. Hunninghake, U of I
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Distribution information can be found at
 http://genome.uiowa.edu/distribution/cgap.html
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES

source
 1..623
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-H-FT2-bj1-i-10-0-UI"
 /issue_type="Aveolar Macrophage"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI CGAP FT2"
 /note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 NCI CGAP FT2 is a subcloned cDNA library constructed from

ORIGIN

Alignment Scores:
 Pred. No.: 1.32e-18 Length: 623
 Score: 32.00 Matches: 32
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 40.51% Indels: 0
 DB: 14 Gaps: 0

US-09-972-032-2 (1-79) x CD742581 (1-623)

QY 48 SerArGrTpProArGrSerAlasArGrTpProTrpSerAlaglyLeuThrValArg 67
 |||||
 Db 622 TCCCGTGGCCCGCAGCGCTCCCGTGGCCCTGGAGTGCAGGTCTTACCGTCCGAGAT 563
 |||||
 QY 68 ArgProGlnLeuGlyGluLeuCysMetGlyArgGly 79
 |||||
 Db 562 CGTCCGCACTGGCGAGCTGTGCATGGCGCGTGGC 527
 |||||

RESULT 10

BQ447041/c
 LOCUS
 DEFINITION UI-H-EUI-bac-p-06-0-UI.s1 NCI CGAP Ctl Homo sapiens cDNA clone
 UI-H-EUI-bac-p-06-0-UI 3', mRNA sequence.
 BQ447041
 ACCESSION
 VERSION
 KEYWORDS EST.
 SOURCE BQ447041.1 GI:21250153
 Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 628)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Dr. Jose Mercuende
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: clone distribution information can be obtained
 from Dr. M. Bento Soares, bento-soares@uiowa.edu
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES

source
 1..628
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-H-EUI-bac-p-06-0-UI"
 /issue_type="Osteoarthritic Cartilage"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI CGAP Ctl"
 /note="Organ: Knee; Vector: pT7T3-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I;
 NCI CGAP Ctl is a normalized cDNA library containing the
 following tissue(s): Osteoarthritic Cartilage The library
 was constructed according to Bonaldo, Lennon and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA
 synthesis was primed with an oligo-dT primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I

a pool of 81 RNA samples from Alveolar Macrophages
 challenged with different treatments. The library was
 subcloned according to Bonaldo, Lennon and Soares, Genome
 Research, 6:791-806, 1996. The tissue was provided by Dr.
 Gary W. Hunninghake of the University of Iowa.
 TAG TISSUE=Human Lung Alveolar Macrophage
 TAG LIB=UI-H-FT2
 TAG_SEQ=GGCATGCGC"

adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dfr)18 tail. The sequence tag for this library is TGATCACGCT.

TAG TISSUE=osteochondritic cartilage
TAG LIB=UI-H-EUI
TAG_SEQ=TGATCACGCT"

ORIGIN

Alignment Scores:

Pred. No.: 1.46e-10 Length: 628
Score: 23.00 Matches: 23
Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 29.11% Indels: 0
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x B0447041 (1-628)

QY 57 TtrProtrpSerAlaGlyLeuThrValArgAspArgProGlnLeuGlyGluLeuCysMet 76
Db 601 TGGCCCTGGAGTGCAGCTCTTACCGTCGAGATCGTCGCACTGGGGGAGCTGTGCATG 542

QY 77 GlyArgGly 79
Db 541 GGGCGTGGC 533

RESULT 11

BM806108 1084 bp mRNA linear EST 05-MAR-2002
LOCUS AGENCOURT_6553891 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5555626
DEFINITION 5', mRNA sequence.

ACCESSION BM806108

VERSION BM806108.1 GI:19122931

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: L1AM12276 row: b column: 11

High quality sequence stop: 672.

Location/Qualifiers

1..1084

FEATURES

source

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5555626"

/tissue_type="leiomyosarcoma"

/lab_host="DH10B (phage-resistant)"

/clone_lib="NIH_MGC_71"

/note="Organ: uterus; Vector: pCMV-SPORT6; Site 1: NotI;

Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.

Average insert size 2.1 kb. "

ORIGIN

Alignment Scores:

Pred. No.: 2.05e-09 Length: 1084
Score: 22.00 Matches: 22

Percent Similarity: 100.00% Conservativeness: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 27.95% Indels: 0
DB: 12 Gaps: 0

US-09-972-032-2 (1-79) x BM806108 (1-1084)

QY 1 MetCysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThr 20
Db 672 ATGTGTGGGAGACCGCTCGGTAAGCGCTTGGATGGCTTCGCTGATGCACATTGGACC 731

QY 21 GlyLeu 22
Db 732 GGGCTC 737

RESULT 12

BF364571/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 282)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,

O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

PUBLISHED

10737800

COMMENT

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?l=PM1&t2=PM1-NN1084-

240900-005-G05&t3=2000-09-24&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 27

High quality sequence stop: 282.

Location/Qualifiers

1..282

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/dev_stage="Adult"

/clone_lib="NN1084"

/note="Organ: nervous normal; Vector: puc18; Site 1: SmaI;

Site 2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

Alignment Scores:

Pred. No.: 1.43e-05 Length: 282
Score: 17.00 Matches: 17
Percent Similarity: 100.00% Conservativeness: 0

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 424)
 REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITLE Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: Stratagene, Inc.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 418.
 FEATURES
 Location/Qualifiers
 1..424
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1085188"
 /tissue_type="gastric tumor"
 /lab_host="SOLR (kanamycin resistant)"
 /clone_lib="NCI CGAP Gas1"
 /note="Organ: stomach; Vector: Bluescript SK-; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. Pooled gastric tumors. 5' adaptor sequence: 5'
 GAATTCGGCACGAG 3' 3' adaptor sequence: 5'
 CTCGAGTTTTTTTTTTTTTT 3' Average insert size: 1.0 kb."

ORIGIN

Alignment Scores: 310 Length: 424
 Pred. No.: 9.00 Matches: 9
 Score: 100.00% Conservative: 0
 Percent Similarity: 100.00% Mismatches: 0
 Best Local Similarity: 11.39% Indels: 0
 Query Match: 9 Gaps: 0
 DB:

US-09-972-032-2 (1-79) x AA593543 (1-424)

QY 23 TrpThrGlyLeuGlyGluGlyGlnGlu 31
 |||||
 Db 374 TGGACAGGGCTCGAGAGGCCAGGAG 400

Search completed: July 13, 2004, 15:53:46
 Job time : 2893 secs


```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: July 12, 2004, 18:17:11 ; Search time 403 Seconds
        (without alignments)
        955.426 Million cell updates/sec

Title:
Perfect score: 456
Sequence: 1 MCGRRVRSAGCGFADAHWT.....SAGLTVRDPQLGELCMGRG 79

Scoring table: BLOSUM62
                Xgapop 10.0 , Xgapext 0.5
                Ygapop 10.0 , Ygapext 0.5
                Fgapop 6.0 , Fgapext 7.0
                Delop  6.0 , Delext  7.0

Searched: 3183909 seqs, 2436941669 residues

Total number of hits satisfying chosen parameters: 6367818

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xilh
-Q=/cgn2_1/USPTO.spool/US09972032/runat_06072004_121455_8495/app_query.fasta_1.263
-DB=published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdd -LIST=45 -DCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09972032@cgn_1_1_511@runat_06072004_121455_8495
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	DB ID	Description
<p>Alignment Scores:</p>				

	1	456	100.0	990	13	US-09-972-032-1	Sequence 1, Appli
1	2	95	20.8	3642	16	US-10-369-493-44624	Sequence 44624, A
C	3	95	20.8	822900	16	US-10-292-798-1393	Sequence 1393, Ap
C	4	94.5	20.7	6395	17	US-10-437-963-34802	Sequence 34802, A
	5	94.5	20.7	24081	15	US-10-132-134-13	Sequence 13, Appli
	6	94.5	20.7	52101	15	US-10-132-134-13	Sequence 1, Appli
C	7	94	20.6	699	14	US-10-006-522-41	Sequence 41, Appli
	8	93.5	20.5	943	13	US-10-245-752-111	Sequence 111, App
	9	93.5	20.5	943	13	US-10-245-752-111	Sequence 111, App
	10	93.5	20.5	943	15	US-10-245-103-111	Sequence 111, App
	11	93.5	20.5	943	15	US-10-245-107-111	Sequence 111, App
	12	93.5	20.5	943	15	US-10-245-143-111	Sequence 111, App
	13	93.5	20.5	943	15	US-10-245-771-111	Sequence 111, App
	14	93.5	20.5	943	15	US-10-245-851-111	Sequence 111, App
	15	93.5	20.5	943	15	US-10-245-883-111	Sequence 111, App
	16	93.5	20.5	943	15	US-10-237-535-111	Sequence 111, App
	17	93.5	20.5	943	15	US-10-238-183-111	Sequence 111, App
	18	93.5	20.5	943	15	US-10-238-283-111	Sequence 111, App
	19	93.5	20.5	943	15	US-10-238-370-111	Sequence 111, App
	20	93.5	20.5	943	15	US-10-245-055-111	Sequence 111, App
	21	93.5	20.5	943	15	US-10-245-147-111	Sequence 111, App
	22	93.5	20.5	943	15	US-10-245-730-111	Sequence 111, App
	23	93.5	20.5	943	15	US-10-245-739-111	Sequence 111, App
	24	93.5	20.5	943	15	US-10-246-310-111	Sequence 111, App
	25	93.5	20.5	943	15	US-10-239-196-111	Sequence 111, App
	26	93.5	20.5	943	15	US-10-243-024-111	Sequence 111, App
	27	93.5	20.5	943	15	US-10-243-409-111	Sequence 111, App
	28	93.5	20.5	943	15	US-10-245-621-111	Sequence 111, App
	29	93.5	20.5	943	15	US-10-245-880-111	Sequence 111, App
	30	93.5	20.5	943	15	US-10-245-033-111	Sequence 111, App
	31	93.5	20.5	943	15	US-10-243-095-111	Sequence 111, App
	32	93.5	20.5	943	15	US-10-245-185-111	Sequence 111, App
	33	93.5	20.5	943	15	US-10-245-427-111	Sequence 111, App
	34	93.5	20.5	943	15	US-10-245-473-111	Sequence 111, App
	35	93.5	20.5	943	15	US-10-245-770-111	Sequence 111, App
	36	93.5	20.5	943	15	US-10-245-877-111	Sequence 111, App
	37	93.5	20.5	943	15	US-10-246-976-111	Sequence 111, App
	38	93.5	20.5	943	15	US-10-243-320-111	Sequence 111, App
	39	93.5	20.5	943	15	US-10-242-743-111	Sequence 111, App
	40	93.5	20.5	943	15	US-10-242-845-111	Sequence 111, App
	41	93.5	20.5	943	15	US-10-237-636-111	Sequence 111, App
	42	93.5	20.5	943	15	US-10-238-325-111	Sequence 111, App
	43	93.5	20.5	943	15	US-10-238-346-111	Sequence 111, App
	44	93.5	20.5	943	15	US-10-238-411-111	Sequence 111, App
	45	93.5	20.5	943	15	US-10-243-124-111	Sequence 111, App

ALIGNMENTS

RESULT 1

US-09-972-032-1

; Sequence 1, Application US/09972032

; Publication No. US20020086361A1

; GENERAL INFORMATION:

; APPLICANT: Case Western Reserve University

; APPLICANT: Montano, Monica

; APPLICANT: Sutton, Amelia

; TITLE OF INVENTION: A Modulator of Antiestrogen Pharmacology

; FILE REFERENCE: 27708/04003

; CURRENT APPLICATION NUMBER: US/09/972,032

; PRIOR FILING DATE: 2001-10-05

; PRIOR APPLICATION NUMBER: US 60/238,190

; PRIOR FILING DATE: 2000-10-05

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 1

; LENGTH: 990

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-972-032-1

```
Pred. No.: 1,22e-36 Length: 990
Score: 456.00 Matches: 79
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x US-09-972-032-1 (1-990)
QY 1 MetCysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThr 20
DB 203 ATGTGTGGAGACCGCGCTGAGCGCTGATGCTGCTGATGACATGGACC 262
QY 21 GlyLeuTrpThrGlyLeuGlyGlnGluGlyGlnGluGlyGlnGluGlnAla 40
DB 263 GGGCTCTGGAGCTGGCTAGGGGAAGGCGAGAGGCGGAATTGGGCGGAGGCCAGGCC 322
QY 41 SerProThrProAspCysAlaSerArgTrpProArgSerAlaSerArgTrpProTrpSer 60
DB 323 TCGCGGACCCCGGACTGCGCTCCCGGTGGCCCGCAGCGCTCCCGGTGGCCCTGGAGT 382
QY 61 AlaGlyLeuThrValArgAspArgProGlnLeuGlyGlnLeuCysMetGlyArgGly 79
DB 383 GCAGGTCTTACCGTCCGAGATCGTCCGAACTGGCGAGCTGTGCATGGGCGCTGGC 439

RESULT 2
US-10-369-493-44624
; Sequence 44624, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 44624
; LENGTH: 3642
; TYPE: DNA
; ORGANISM: SYNECHOCOCCUS SP. WH 8102
US-10-369-493-44624

Alignment Scores:
Pred. No.: 2,92 Length: 3642
Score: 95.00 Matches: 31
Percent Similarity: 39.56% Conservative: 5
Best Local Similarity: 34.07% Mismatches: 33
Query Match: 20.83% Indels: 22
DB: 16 Gaps: 6

US-09-972-032-2 (1-79) x US-10-369-493-44624 (1-3642)
QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThr--- 20
DB 2820 TGCCGATCTCTGGGCGACCTCGGTGGATGGTGGTGGGATCTCGGATTCGGTTGTCG 2879
QY 21 GlyLeuTrpThrGlyLeuGlyGlu-----GlyGlnGluGly----- 32
DB 2880 GGATGCTTCCCGCAGCTGGTGGTGGTGAACACAGGACACACAGATGGTGGCCTTGCT 2939
QY 33 -----GlyLeuGlyProGluGlyGlnAlaSerProThr-----ProaspCys 46
DB 2940 GGAGGAACAGAGGAGATCAACCTCTGGCCAGCCTCACCCTGCTGTGATGGCCCAAGG 2999
QY 47 AlaSerArgTrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThrValArg 66
DB 47374
```

```
Db 3000 ACGAATCTATGATCCGCCCGCCCTATGGCGCTTGGTTGCAGGCCTT----- 3050
QY 67 AspArgProGlnLeuGlyGlnLeuCysMetGly 77
DB 3051 GATCGA-----CAGTGGTGCATGGGA 3071

RESULT 3
US-10-292-798-1393/c
; Sequence 1393, Application US/10292798
; Publication No. US20030235833A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAL, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABRATANI, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 08435/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1393
; LENGTH: 822900
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(822900)
; NAME/KEY: CDS
; LOCATION: (201)..(1068)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (140545)..(140693)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (261786)..(261845)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (273663)..(273702)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (347633)..(347711)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (482589)..(482596)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (534176)..(534210)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (822485)..(822700)
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (4848)..(4947)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (4966)..(4966)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (17785)..(17884)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (61159)..(61258)
; OTHER INFORMATION: a, t, c, g, unknown or other
```

```

; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (67605)..(67704)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (74625)..(74724)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (85854)..(85953)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (367573)..(367573)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (367588)..(367588)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (367595)..(367595)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (367613)..(367614)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (367985)..(367985)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (367993)..(367993)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (367995)..(367996)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (367998)..(367998)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (370273)..(370277)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (370279)..(370279)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (370281)..(370282)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (370285)..(370287)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (370289)..(370291)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (385643)..(385742)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (409961)..(410060)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:

```

```

; NAME/KEY: modified_base
; LOCATION: (410096)..(410096)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (417384)..(417483)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (724960)..(725059)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (726106)..(726205)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (727470)..(727569)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (728849)..(728948)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (730296)..(730395)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (731863)..(731962)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (734124)..(734223)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (734441)..(734441)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (735752)..(735851)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (738576)..(738675)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (740924)..(741023)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (744360)..(744459)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (748430)..(748529)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (754323)..(754422)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:

```

```

Alignment Scores:
Pred. No.: 348
Score: 95.00
Percent Similarity: 37.39%
Best Local Similarity: 32.17%
Query Match: 20.83%
DB: 16
Length: 822900
Matches: 37
Conservative: 6
Mismatch: 29
Indels: 43
Gaps: 5

```

```

US-09-972-032-2 (1-79) x US-10-292-798-1393 (1-822900)

```

QY 3 GlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeu 22
DB 107177 GCGGCACCTGTCGGGGAGACTTGGCTGC-----GCAGAGCCACCGAGTGGG--- 107127
QY 23 TrpThrGlyLeuGlyGluGlyGlnGluGly----- 32
DB 107126 -----GGAGGAGGCTTAGGCATGCGGGCTGTGGTCCACAGCCCTGCTACAGGGAA 107073
QY 32 ----- 32
DB 107072 GCAGCTAAGGCTGTGAGAAATCGACGACAGCAGCTGCTGGCCAGGCTGCTAAGCCCT 107013
QY 33 ---GlyIleGlyProGluGlyGlnAlaSerProThrProAspCysAlaSerArgTrpPro 51
DB 107012 CACTGCTTGGGGCAGCGGGCCCGCCAGCCACGCCACCCAGAACTCGCGTGGCCC 106953
QY 52 ArgSerAlaSerArgTrpProTrpSerAla---GlyLeuThrValArgAspArgProGln 70
DB 106952 ACAAGCACCGCGCAGCCCTGGTCTGCTGCGCCACACCTCTCCCTCCACACCTCCCGCAA 106893
QY 71 -LeuGlyGluLeuCysMet-----GlyArgGly 79
DB 106892 GCTGAGGAGGCTGCTCTAGCCTCGCCGACCCAGGAGGGGC 106850

RESULT 4

US-10-437-963-34802/c
; Sequence 34802, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 34802
; LENGTH: 6395
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_38782C.1
US-10-437-963-34802

Alignment Scores:
Pred. No.: 5.38 Length: 6395
Score: 94.50 Matches: 31
Percent Similarity: 43.37% Conservative: 5
Best Local Similarity: 37.35% Mismatches: 24
Query Match: 20.72% Indels: 23
DB: 17 Gaps: 6

US-09-972-032-2 (1-79) x US-10-437-963-34802 (1-6395)

QY 5 ProArgArgValSerAlaGly---CysGlyPheAlaAspAla---HisTrpThrGlyLeu 22
DB 5977 CTTCTGCTCAGTGGCGGATCTCTGTTTCAGAAAGATGCACCACTGGACTGGGTTA 5918
QY 23 -----TrpThrGlyLeuGlyGluGly 29
DB 5917 GGAGGGTGGGATCTGCTGTTTCAGGAGGTGCACCACTGGACTGGGTTAGGAGAGGT 5858
QY 30 GlnGluGlyGlyIleGlyProGlu-----GlyGlnAlaSerProThrPro 44
DB 5857 TCAGTGCAGGGCTTGTGACGCATCAGCCATTGGCTCGGTGACGCATCTCTCCACCA 5798

QY 45 AspCysAlaSer---ArgTrpProArgSer-----AlaSerArgTrpProTrpSerAla 61
DB 5797 GATTGGGTACAGGGCTTTGGTCTGATTCATGGTACCATCCTTCTCCAGTTTCTTCT 5738
QY 62 GlyLeuThr 64
DB 5737 GGCATGTCA 5729

RESULT 5

US-10-132-134-13
; Sequence 13, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
; FILE REFERENCE: 3012-2US
; CURRENT APPLICATION NUMBER: US/10/132,134
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 24081
; TYPE: DNA
; ORGANISM: Streptomyces platensis subsp. roseaceus
US-10-132-134-13

Alignment Scores:

Pred. No.: 17.3 Length: 24081
Score: 94.50 Matches: 27
Percent Similarity: 42.42% Conservative: 1
Best Local Similarity: 40.91% Mismatches: 30
Query Match: 20.72% Indels: 8
DB: 15 Gaps: 3

US-09-972-032-2 (1-79) x US-10-132-134-13 (1-24081)

QY 9 SerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeuTrpThrGlyLeuGly-G1 28
DB 12490 TCGCGGCTCGGCTTCGAGGACGCGACCCGCCCTTCGCCGAGGCGCTGGACTGC 12549
QY 28 uGlyGlnGluGlyIleGlyProGluGlyGlnAlaSerProThrProAspCysAlaSe 48
DB 12550 CCGGTCAAGTCTGTCACCTGGGGCTACTGGGGCAACGTCGCCGACGAGCTCTCGCGCAC 12609
QY 48 r-----ArgTrpProArgSer-----AlaSerArgTrpPro---TrpSerAl 61
DB 12610 GTGACGTCATGGGACTGGCCCCGATCGCCCCGCCACGCGCATGGCGCAGCTGGAGCGC 12669
QY 61 aGlyLeuThrValArg 66
DB 12670 CTCCTGGCGGCGCCGC 12685

RESULT 6

US-10-132-134-1
; Sequence 1, Application US/10132134
; Publication No. US20030171562A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Yang, Xianshu
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES
; FILE REFERENCE: 3012-2US
; CURRENT APPLICATION NUMBER: US/10/132,134
; CURRENT FILING DATE: 2002-04-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 52101
; TYPE: DNA

```

; ORGANISM: Streptomyces platensis subsp. roseaceus
US-10-132-134-1

Alignment Scores:
Pred. No.: 34.2 Length: 52101
Score: 94.50 Matches: 27
Percent Similarity: 42.42% Conservative: 1
Best Local Similarity: 40.91% Mismatches: 30
Query Match: 20.72% Indels: 8
DB: 15 Gaps: 3

US-09-972-032-2 (1-79) x US-10-132-134-1 (1-52101)

Qy 9 SexAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeuTrpThrGlyLeuGly-GI 28
Db 30123 TCGGCGCCCTGCGCCCTTCGAGGACGCGCAGCGCCGCTGCGGAGGCGCTGGACTGC 30182

Qy 28 uGlyGlnGluGlyGlyLeuGlyProGluGlyGlnAlaSerProThrProAspCysAla 48
Db 30183 CGGGTCAAGGTCTGCAACTGGGGCTACTGGGGCAACGTCCCGCAGCAGCTCTCTGCGCGAC 30242

Qy 48 r-----ArgTrpProArgSer-----AlaSerArgTrpPro---TrpSerAl 61
Db 30243 GTGACGTCCATGGAGTGGCCCGGATCGCCCGCGCACGCGCGATGGCGCACTGGAGCGC 30302

Qy 61 aGlyLeuThrValArg 66
Db 30303 CTCCTGCGCGCGCGC 30318

RESULT 7
US-10-006-922-41/c
; Sequence 41, Application US/10006922
; Publication No. US20020197676A1
; GENERAL INFORMATION:
; APPLICANT: Lukyanov, Sergey A
; APPLICANT: Pradkov, Arcady F.
; APPLICANT: Labas, Yulii A.
; APPLICANT: Matz, Mikhail V.
; APPLICANT: Tetsikh, Alexey
; TITLE OF INVENTION: No. US20020197676A1el Chromophores/Fluorophores and
; FILE REFERENCE: CLON-035CIP
; CURRENT APPLICATION NUMBER: US/10/006,922
; CURRENT FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 09/120,330
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/457,898
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,144
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/458,477
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/457,556
; PRIOR FILING DATE: 1999-12-09
; PRIOR APPLICATION NUMBER: 09/444,338
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 41
; TYPE: DNA
; ORGANISM: Anemonia sulcata
US-10-006-922-41

Alignment Scores:
Pred. No.: 0.856 Length: 699
Score: 94.00 Matches: 33
Percent Similarity: 43.02% Conservative: 4
Best Local Similarity: 38.37% Mismatches: 31
Query Match: 20.61% Indels: 18
DB: 14 Gaps: 6

US-09-972-032-2 (1-79) x US-10-006-922-41 (1-699)

; ORGANISM: Streptomyces platensis subsp. roseaceus
US-10-132-134-1

Alignment Scores:
Pred. No.: 34.2 Length: 52101
Score: 94.50 Matches: 27
Percent Similarity: 42.42% Conservative: 1
Best Local Similarity: 40.91% Mismatches: 30
Query Match: 20.72% Indels: 8
DB: 15 Gaps: 3

US-09-972-032-2 (1-79) x US-10-132-134-1 (1-52101)

Qy 9 SexAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeuTrpThrGlyLeuGly-GI 28
Db 30123 TCGGCGCCCTGCGCCCTTCGAGGACGCGCAGCGCCGCTGCGGAGGCGCTGGACTGC 30182

Qy 28 uGlyGlnGluGlyGlyLeuGlyProGluGlyGlnAlaSerProThrProAspCysAla 48
Db 30183 CGGGTCAAGGTCTGCAACTGGGGCTACTGGGGCAACGTCCCGCAGCAGCTCTCTGCGCGAC 30242

Qy 48 r-----ArgTrpProArgSer-----AlaSerArgTrpPro---TrpSerAl 61
Db 30243 GTGACGTCCATGGAGTGGCCCGGATCGCCCGCGCACGCGCGATGGCGCACTGGAGCGC 30302

Qy 61 aGlyLeuThrValArg 66
Db 30303 CTCCTGCGCGCGCGC 30318

RESULT 8
US-10-245-752-111
; Sequence 111, Application US/10245752
; Publication No. US20030064473A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630RIC66
; CURRENT APPLICATION NUMBER: US/10/245,752
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 111
; LENGTH: 943
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-245-752-111

Alignment Scores:
Pred. No.: 1.25 Length: 943

```

Score: 93.50 Matches: 36
 Percent Similarity: 35.25% Conservative: 7
 Best Local Similarity: 29.51% Mismatches: 31
 Query Match: 20.50% Indels: 48
 DB: 13 Gaps: 6

US-09-972-032-2 (1-79) x US-10-245-752-111 (1-943)

QY 2 CysGlyArgProArgArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
 Db 138 TGTCCAGACGCTGGAGATGGCGCAGCGCAGCGCTGTGGAGTCTCTGCTGGC 197
 QY 22 LeuTrpThrGlyLeuGlyGlnGluGlyGlyLeuGlyProGlu----- 37
 Db 198 TGTGTGACAGACCGCGGAGGCGCA-----GCCCTGGGGCCAGAGCCGCGCAGGTGG 251
 QY 38 -----GlyGlnAlaSerProThrPro 44
 Db 252 ACGCACATGACTGTGAGCGCTGGGCTCCGAGGAGCGCGCTCCAGAGTCCC 311
 QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52
 Db 312 GAGCAGCGTCAAACTGCAGTTCACATGATGCGCGCTCCAGAGTCCC 371
 QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66
 Db 372 CGCTCAGCGCAGGCGCAGCTCACCTTCTCTGGGGCTGGTGGCCCTGCCCTGCTGTAC 431
 QY 66 gAspArgProGlnLeu-----GlyGlnLeuCysMe 76
 Db 432 CCGACGCGCCCGTGTCTGGGAGGAGCCATGCGCGCTGCATTCACAGTGGCGAGTTTGTCC 491
 QY 76 tGly 77
 Db 492 TGGT 495

RESULT 9

US-10-245-859-111
 ; Sequence 111, Application US/10245859
 ; Publication No. US20030064474A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin
 ; APPLICANT: Eaton, Dan
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Grimaldi, J. Christopher
 ; APPLICANT: Gurney, Austin
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stephan, Jean-Phillippe
 ; APPLICANT: Watanabe, Colin
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin
 ; APPLICANT: Fong, Sherman
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3630R1C78
 ; CURRENT APPLICATION NUMBER: US/10/245, 859
 ; CURRENT FILING DATE: 2002-09-16
 ; PRIOR APPLICATION NUMBER: 10/197942
 ; PRIOR FILING DATE: 2002-07-18
 ; PRIOR APPLICATION NUMBER: 60/059114
 ; PRIOR FILING DATE: 1997-09-17
 ; PRIOR APPLICATION NUMBER: 60/063046
 ; PRIOR FILING DATE: 1997-10-24
 ; PRIOR APPLICATION NUMBER: 60/065027
 ; PRIOR FILING DATE: 1997-11-10
 ; PRIOR APPLICATION NUMBER: 60/079689
 ; PRIOR FILING DATE: 1998-03-27
 ; PRIOR APPLICATION NUMBER: 60/086478
 ; PRIOR FILING DATE: 1998-05-22
 ; PRIOR APPLICATION NUMBER: 60/087607
 ; PRIOR FILING DATE: 1998-06-02
 ; PRIOR APPLICATION NUMBER: 60/089801

; PRIOR FILING DATE: 1998-06-18
 ; PRIOR APPLICATION NUMBER: 60/090557
 ; PRIOR FILING DATE: 1998-06-24
 ; PRIOR APPLICATION NUMBER: 60/090689
 ; PRIOR FILING DATE: 1998-06-25
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 116
 ; SEQ ID NO 111
 ; LENGTH: 943
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 ; US-10-245-859-111

Alignment Scores:
 Pred. No.: 1.25 Length: 943
 Score: 93.50 Matches: 36
 Percent Similarity: 35.25% Conservative: 7
 Best Local Similarity: 29.51% Mismatches: 31
 Query Match: 20.50% Indels: 48
 DB: 13 Gaps: 6

US-09-972-032-2 (1-79) x US-10-245-859-111 (1-943)

QY 2 CysGlyArgProArgArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
 Db 138 TGTCCAGACGCTGGAGATGGCGCAGCGCAGCGCTGTGGAGTCTCTGCTGGC 197
 QY 22 LeuTrpThrGlyLeuGlyGlnGluGlyGlyLeuGlyProGlu----- 37
 Db 198 TGTGTGACAGACCGCGGAGGCGCA-----GCCCTGGGGCCAGAGCCGCGCAGGTGG 251
 QY 38 -----GlyGlnAlaSerProThrPro 44
 Db 252 ACGCACATGACTGTGAGCGCTGGGCTCCGAGGAGCGCGCTCCAGAGTCCC 311
 QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52
 Db 312 GAGCAGCGTCAAACTGCAGTTCACATGATGCGCGCTCCAGAGTCCC 371
 QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66
 Db 372 CGCTCAGCGCAGGCGCAGCTCACCTTCTCTGGGGCTGGTGGCCCTGCCCTGCTGTAC 431
 QY 66 gAspArgProGlnLeu-----GlyGlnLeuCysMe 76
 Db 432 CCGACGCGCCCGTGTCTGGGAGGAGCCATGCGCGCTGCATTCACAGTGGCGAGTTTGTCC 491
 QY 76 tGly 77
 Db 492 TGGT 495

RESULT 10

US-10-245-103-111
 ; Sequence 111, Application US/10245103
 ; Publication No. US20030068778A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin
 ; APPLICANT: Eaton, Dan
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Grimaldi, J. Christopher
 ; APPLICANT: Gurney, Austin
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stephan, Jean-Phillippe
 ; APPLICANT: Watanabe, Colin
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin
 ; APPLICANT: Fong, Sherman
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3630R1C112
 ; CURRENT APPLICATION NUMBER: US/10/245, 103
 ; CURRENT FILING DATE: 2002-09-17


```
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 111
; LENGTH: 943
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-245-103-111
Alignment Scores:
Pred. No.: 1.25 Length: 943
Score: 93.50 Matches: 36
Percent Similarity: 35.25% Conservative: 7
Best Local Similarity: 29.51% Mismatches: 31
Query Match: 20.50% Indels: 48
DB: 15 Gaps: 6
US-09-972-032-2 (1-79) x US-10-245-103-111 (1-943)
QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
Db 138 TGTGCCAGACGCTGGAGGATGGCGCAGCGAGCTGGGGCTGTGGAGTCTCTGCTGCG 197
QY 22 LeuTrpThrGlyLeuGlyGluGlyGlnGluGlyGlyLeuGlyGlyProGlu----- 37
Db 198 TGTGGACAGACCCCGGGAGGCGCA-----GCCCTGGGGCCAGAGCCGCCAGGTGG 251
QY 38 -----GlyGlnAlaSerProThrPro 44
Db 252 ACGCACATGACTGTGAGCGCTGGGGCTCCGAGGCGAGCGCGCTTCACGAGTCCC 311
QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52
Db 312 GAGGACCGCTCAAACTGCAGTTCGACATGATGCGCGCTGCAACCTGGTGGCCAGCGCG 371
QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66
Db 372 CGCTCACCAGCGAGCCAGTCCATTCCTCTCTGGGGCTGGTGGCGCTGCTGTCTAC 431
QY 66 gApArgProGlnLeu-----GlyGlnLeuCysMe 76
Db 432 CCGACGCGCGTGTCTGGGAGGAGCCATGCCCGCTGCATTCCAACTGGCGAGTTTGTCC 491
QY 76 tGly 77
Db 492 TGGT 495
RESULT 11
US-10-245-107-111
; Sequence 111, Application US/10245107
; Publication No. US20030068779A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Baton, Dan
```

```
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630RIC71
; CURRENT APPLICATION NUMBER: US/10/245.107
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 111
; LENGTH: 943
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-245-107-111
Alignment Scores:
Pred. No.: 1.25 Length: 943
Score: 93.50 Matches: 36
Percent Similarity: 35.25% Conservative: 7
Best Local Similarity: 29.51% Mismatches: 31
Query Match: 20.50% Indels: 48
DB: 15 Gaps: 6
US-09-972-032-2 (1-79) x US-10-245-107-111 (1-943)
QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
Db 138 TGTGCCAGACGCTGGAGGATGGCGCAGCGAGCTGGGGCTGTGGAGTCTCTGCTGCG 197
QY 22 LeuTrpThrGlyLeuGlyGluGlyGlnGluGlyGlyLeuGlyGlyProGlu----- 37
Db 198 TGTGGACAGACCCCGGGAGGCGCA-----GCCCTGGGGCCAGAGCCGCCAGGTGG 251
QY 38 -----GlyGlnAlaSerProThrPro 44
Db 252 ACGCACATGACTGTGAGCGCTGGGGCTCCGAGGCGAGCGCGCTTCACGAGTCCC 311
QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52
Db 312 GAGGACCGCTCAAACTGCAGTTCGACATGATGCGCGCTGCAACCTGGTGGCCAGCGCG 371
QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66
Db 372 CGCTCACCAGCGAGCCAGTCCATTCCTCTCTGGGGCTGGTGGCGCTGCTGTCTAC 431
```

```
QY 66 gAspArgProGlnLeu-----GlyGlnLeuCysMe 76
Db 432 CCAGCGCCCGTGTGGGAGGAGGCATGGCCGTCATTCCAACTGGCGAGTTTGTCC 491
QY 76 tGly 77
Db 492 TGGT 495

RESULT 12
US-10-245-143-111
; Sequence 111, Application US/10245143
; Publication No. US20030068780A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C90
; CURRENT APPLICATION NUMBER: US/10/245,143
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 111
; LENGTH: 943
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-245-143-111

Alignment Scores:
Pred. No.: 1,25 Length: 943
Score: 93,50 Matches: 36
Percent Similarity: 35,25% Conservatives: 7
Best Local Similarity: 29,51% Mismatches: 31
Query Match: 20,50% Indels: 48
DB: 15 Gaps: 6

US-09-972-032-2 (1-79) x US-10-245-143-111 (1-943)

QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
Db 138 TGTGCCAGACGCTGGAGAGATGGCGGCGAGCGCATGGCGGTGGAGGTCTCTGTGTC 197
QY 22 LeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlyGlyGlyProGlu----- 37
```

```
Db 198 TGGTGACAGGACCCGGGGAGGGCCGA-----CCCTGGGGCCAGAGCCGGCCAGGTGG 251
QY 38 -----GlyGlnAlaSerProThrPro 44
Db 252 ACGCACATGACTGTGAGCGCTGGGCTGGGGCTCGAGGAGGAGCGGCTTCAGAGAGTCCC 311
QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52
Db 312 GAGGCACCGTCAAACTGCAGTTCGACATGATGCGCGCTGCACACTGGTGGCCACGCGCG 371
QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66
Db 372 CGTCCACCGGAGCGGAGCTCACCTTCCTCTGGGGCTGGTGGGCTGCCCTGCTGTGTC 431
QY 66 gAspArgProGlnLeu-----GlyGlnLeuCysMe 76
Db 432 CCAGCGCCCGTGTGGGAGGAGGCATGGCCGTCATTCCAACTGGCGAGTTTGTCC 491
QY 76 tGly 77
Db 492 TGGT 495

RESULT 13
US-10-245-771-111
; Sequence 111, Application US/10245771
; Publication No. US20030068781A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3630R1C98
; CURRENT APPLICATION NUMBER: US/10/245,771
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 111
; LENGTH: 943
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-245-771-111
```

Alignment Scores:

```
Pred. No.: 1.25 Length: 943
Score: 93.50 Matches: 36
Percent Similarity: 35.25% Conservative: 7
Best Local Similarity: 29.51% Mismatches: 31
Query Match: 20.50% Indels: 48
DB: 15 Gaps: 6

US-09-972-032-2 (1-79) x US-10-245-771-111 (1-943)

QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
Db 138 TGTCCAGACGCTGAGAGTGGCGCAGGCGCAGCGCTGGGGCTGTGAGGTCTCTGTCGC 197
QY 22 LeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlyGlyGlyGlyGlyGlyGlyGly 37
Db 198 TGTGTGACAGACCGCGGAGGCGCGA-----GCCCTGGGGCCAGAGCGCGCGTGG 251
QY 38 -----GlyGlnAlaSerPro 44
Db 252 ACGCACATGACTGTGAGCGCTGGGGCTCCGAGGAGCGCGGCTTCCAGAGTCCC 311
QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52
Db 312 GAGGACCGCTCAAACTGCAGTTGCATGATGCGCGCTGCAACCTGGTGGCGCACGCGCG 371
QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValar 66
Db 372 CGTCAACGAGCGACCTACCTTCTCTGGGGCTGGTGGCGCTGCCCTCTGTCTCAC 431
QY 66 GAspArgProGlnLeu-----GlyGluLeuCysMe 76
Db 432 CCGAGCGCGCGTGTGGAGGAGCCATGCGCGCTGCATTCCAACTGGCGAGTTTGTCC 491
QY 76 tGly 77
Db 492 TGGT 495
```

RESULT 14

```
US-10-245-851-111
; Sequence 111, Application US/10245851
; Publication No. US20030068782A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3630R1C93
; CURRENT APPLICATION NUMBER: US/10/245,851
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
```

```
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 111
; LENGTH: 943
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-245-851-111
```

Alignment Scores:

```
Pred. No.: 1.25 Length: 943
Score: 93.50 Matches: 36
Percent Similarity: 35.25% Conservative: 7
Best Local Similarity: 29.51% Mismatches: 31
Query Match: 20.50% Indels: 48
DB: 15 Gaps: 6

US-09-972-032-2 (1-79) x US-10-245-851-111 (1-943)

QY 2 CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
Db 138 TGTCCAGACGCTGAGAGTGGCGCAGGCGCAGCGCTGGGGCTGTGAGGTCTCTGTCGC 197
QY 22 LeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlyGlyGlyGlyGlyGlyGlyGly 37
Db 198 TGTGTGACAGACCGCGGAGGCGCGA-----GCCCTGGGGCCAGAGCGCGCGTGG 251
QY 38 -----GlyGlnAlaSerPro 44
Db 252 ACGCACATGACTGTGAGCGCTGGGGCTCCGAGGAGCGCGGCTTCCAGAGTCCC 311
QY 45 Asp-----CysAlaSerArg-----TrpProArg--- 52
Db 312 GAGGACCGCTCAAACTGCAGTTGCATGATGCGCGCTGCAACCTGGTGGCGCACGCGCG 371
QY 53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValar 66
Db 372 CGTCAACGAGCGACCTACCTTCTCTGGGGCTGGTGGCGCTGCCCTCTGTCTCAC 431
QY 66 GAspArgProGlnLeu-----GlyGluLeuCysMe 76
Db 432 CCGAGCGCGCGTGTGGAGGAGCCATGCGCGCTGCATTCCAACTGGCGAGTTTGTCC 491
QY 76 tGly 77
Db 492 TGGT 495
```

RESULT 15

```
US-10-245-883-111
; Sequence 111, Application US/10245883
; Publication No. US20030068783A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin
; APPLICANT: Eaton, Dan
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Goddard, Audrey
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Phillippe
; APPLICANT: Watanabe, Colin
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; APPLICANT: Fong, Sherman
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3630R1C70
```

```

; CURRENT APPLICATION NUMBER: US/10/245,883
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: 10/197942
; PRIOR FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: 60/059114
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/063046
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/065027
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: 60/079689
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/086478
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/089801
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090689
; PRIOR FILING DATE: 1998-06-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 111
; LENGTH: 943
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-245-883-111

Alignment Scores:
Pred. No.:      1.25      Length:      943
Score:          93.50     Matches:     36
Percent Similarity: 35.25% Conservatives: 7
Best Local Similarity: 29.51% Mismatches:    31
Query Match:      20.50% Indels:       48
DB:               15      Gaps:        6

US-09-972-032-2 (1-79) x US-10-245-883-111 (1-943)

Qy      2 CysGlyArgProArgArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly 21
Db      138 TGTGCCAGACGCTGGAGGATGGCGCAGCGCGAGCGTGGGGCTGTGGAGTCTCTGCTGCG 197
Qy      22 LeuTrpThrGlyLeuGlyGluGlyGlnGluGlyGlyGlyProGlu----- 37
Db      198 TGTGGACAGACCCGCGGAGGCCGA-----GCCCTGGGGCCAGAGCGCGCCAGGTGG 251
Qy      38 -----GlyGlnAlaSerProThrPro 44
Db      252 ACGCACATGACTGTAGCGCGCTGGGCTGGGCTCCGAGGCGAGCGGCTTCAGAGTCC 311
Qy      45 Asp-----CysAlaSerArg-----TrpProArg--- 52
Db      312 GAGGCACCGTCAAACTCGAGTTCGACATGATGCGCGCTGCAACCTGGTGGCCACGCGCG 371
Qy      53 -----SerAlaSerArgTrpProTrpSerAla-GlyLeuThrValAr 66
Db      372 CGCTCACCAGCGCAGCTCACCTCTCTCTGGGGCTGTGGGCTGCCCCCTGCTGTAC 431
Qy      66 gAspArgProGlnLeu-----GlyGlnLeuCysMe 76
Db      432 CCGACGCCCCGCTGCTGGGAGAGGCCCATGCGCGCTGCATTCCAACTGGGCGAGTTTGTCC 491
Qy      76 tGly 77
Db      492 TGGT 495

```


Pred. No.:	2,11e+05	Length:	4411529
Score:	86.50	Matches:	28
Percent Similarity:	39.33%	Conservative:	7
Best Local Similarity:	31.46%	Mismatches:	20
Query Match:	18.97%	Indels:	35
DB:	3	Gaps:	6

US-09-972-032-2 (1-79) x US-09-103-840A-1 (1-4411529)

Qy	2	CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly	21
Dd	2199603	TGGCGTCAGCCGGTGG-AAT-----TGC GGTTTCTCG-----	2199634
Qy	22	LeuTrpThrGlyLeuGlyGluGlyGlnGluGlyVilegylProGluGlyGlnAlaSer	41
Dd	2199635	TTTGGG---GGCACCGGTGTCGTAGCAGCGTGCATTGGG-----GGTTCGTGAAG	2199685
Qy	42	ProThr-----	43
Dd	2199686	CCGACGGTGTAAGGGTGGCGTAGCGTGGCTATTTTGGCGGCCACTTCGTGTTTCGACT	2199745
Qy	44	--ProAspCysAlaSerArgTrp-----ProArgSerAlaSerArgTrpProTrp	59
Dd	2199746	CCGATGAGTTGTGTCCGGCGTTGGCGGATATCCC CGCCAATGCTTGTGCTTGGCGCTGG	2199805
Qy	60	SerAlaGlyLeuThrValArgAspArg	68
Dd	2199806	CGAGTCGTTGTTCCGGCAAGTGGCG	2199832

RESULT 12

US-09-773-816-1
; Sequence 1, Application US/09773816
; Patent No. 6340774
; GENERAL INFORMATION:
; APPLICANT: Stanford University
; APPLICANT: Khosla, Chaitan
; TITLE OF INVENTION: NON-STEROIDAL ESTROGEN-RECEPTOR
; TITLE OF INVENTION: ANTAGONISTS
; FILE REFERENCE: 28600-20210.00
; CURRENT APPLICATION NUMBER: US/09/773,816
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/243,458
; PRIOR FILING DATE: 2000-10-25
; PRIOR APPLICATION NUMBER: US 60/179,305
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 23673
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(23623)
; OTHER INFORMATION: n = A,T,C or G
US-09-773-816-1

Alignment Scores:

Pred. No.:	717	Length:	23673
Score:	86.00	Matches:	23
Percent Similarity:	41.27%	Conservative:	3
Best Local Similarity:	36.51%	Mismatches:	23
Query Match:	18.86%	Indels:	14
DB:	4	Gaps:	3

US-09-972-032-2 (1-79) x US-09-773-816-1 (1-23673)

Qy	6	ArgArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeuTrpThrGly	25
Dd	18526	CGAGAACCCTCCCGCGGTGGCCCGGACCTGGAGTTCCITGGGCGAGGTGTGACCGGC	18585
Qy	26	LeuGlyGluGlyGlnGluGlyVilegyl-----ProGluGlyGlnAlaSer	41

Db 18586 TCGGCTCGGCGCGTCTGGGCAACCGCCCGCGAGTTGTCGGCGGCGCAGCAGC 18645

QY 42 ProThrProAspCysAlaSerArgTrpProArgSerAlaSerArgTrpProTrpSerAla 61
 Db 18646 -----GGGTCCGCGTGG-----CGCGCGCCCTGGTCAGCA 18675

QY 62 GlyLeuThr 64
 Db 18676 GGCCGGACA 18684

RESULT 13

US-08-753-247-13
 ; Sequence 13, Application US/08753247
 ; Patent No. 6210929
 ; GENERAL INFORMATION:
 ; APPLICANT: SCHLOKAT, Uwe
 ; APPLICANT: FISCHER, Bernhard
 ; APPLICANT: FALKNER, Falko-Guenther
 ; APPLICANT: DORNER, Friedrich
 ; APPLICANT: EIBL, Johann
 ; TITLE OF INVENTION: A FUSION PROTEIN COMPRISING A FURIN
 ; TITLE OF INVENTION: DERIVATIVE OR A DERIVATIVE OF A FURIN ANALOGUE AND A
 ; NUMBER OF SEQUENCES: 29
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA: US/08/753,247
 ; APPLICATION NUMBER: 29
 ; FILING DATE: 22-NOV-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION NUMBER: AT 1928/95
 ; FILING DATE: 24-NOV-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, Stephen A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 40433/149
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)672-5300
 ; TELEFAX: (202)672-5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ ID NO: 13:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1758 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 1..1755
 ; FEATURE:
 ; NAME/KEY: mat_peptide
 ; LOCATION: 1..1755
 ; US-08-753-247-13

Alignment Scores:

Pred. No.:	47.5	Length:	1758
Score:	85.00	Matches:	30
Percent Similarity:	34.69%	Conservative:	4
Best Local Similarity:	30.61%	Mismatches:	34
Query Match:	18.64%	Indels:	30
DB:	3	Gaps:	6

US-09-972-032-2 (1-79) x US-08-753-247-13 (1-1758)

QY 3 GlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeu 22
 Db 615 GGCCAAACACGGTGTCTGTGGTGTAGGTGTGGCTACAAACCGCCGCTTGGAGGGTGGC 674

QY 23 TrpThrGly----LeuGlyGluGlyGlnGluGly-----IleGlyProGluGlyGln 39
 Db 675 CATGCTGGATGGCGAGGTGCACAGATGCAGTGCAGGCACGCTCGCTGGGCTTGAACCCCAA 734

QY 40 AlaSerProThrProAspCysAla-----SerArgTrp 50
 Db 735 CCACATCCACATCTACATGCGCAGCTGGGGCCCCGAGGATGACGCGACAGACAGTGGATGG 794

QY 51 -----ProArgSerAlaSerArgTrpProTrp-----SerAla 61
 Db 795 GCCAGCCCGCCTCGCGAGGAGGCTTCTTCCTGGGGTTAGCCAGGCGCCGAGGGGGCT 854

QY 62 GlyLeuThrValArgAspArgProGlnLeuGlyGluLeuCysMetGlyArgGly 79
 Db 855 GGGCTC-----CATCTTTGCTGGGCTCGGG 881

RESULT 14

US-08-753-247-17
 ; Sequence 17, Application US/08753247
 ; Patent No. 6210929
 ; GENERAL INFORMATION:
 ; APPLICANT: SCHLOKAT, Uwe
 ; APPLICANT: FISCHER, Bernhard
 ; APPLICANT: FALKNER, Falko-Guenther
 ; APPLICANT: DORNER, Friedrich
 ; APPLICANT: EIBL, Johann
 ; TITLE OF INVENTION: A FUSION PROTEIN COMPRISING A FURIN
 ; TITLE OF INVENTION: DERIVATIVE OR A DERIVATIVE OF A FURIN ANALOGUE AND A
 ; NUMBER OF SEQUENCES: 29
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA: US/08/753,247
 ; APPLICATION NUMBER: 29
 ; FILING DATE: 22-NOV-1996
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION NUMBER: AT 1928/95
 ; FILING DATE: 24-NOV-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, Stephen A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 40433/149
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)672-5300
 ; TELEFAX: (202)672-5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ ID NO: 17:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1776 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; FEATURE:
 ; NAME/KEY: CDS

```
; LOCATION: 1..1773
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..1773
US-08-753-247-17
Alignment Scores:
Pred. No.: 48 Length: 1776
Score: 85.00 Matches: 30
Percent Similarity: 34.69% Conservative: 4
Best Local Similarity: 30.61% Mismatches: 34
Query Match: 18.64% Indels: 30
DB: 3 Gaps: 6
US-09-972-032-2 (1-79) x US-08-753-247-17 (1-1776)
QY 3 GlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeu 22
Db 615 GGCCAAACACGGTGTCTGTGTAGTGTGGCTTACAGCCCGCATTTGGAGGGGTGG 674
QY 23 TrpThrGly---LeuGlyGluGlyGlnGluGlyGly-----IleGlyProGluGlyGln 39
Db 675 CATGCTGGATGGCGAGGTGCACAGTGCAGTGGAGGACGGTCTGGCTGAACCCCAA 734
QY 40 AlaSerProThrProAspCysAla-----SerArgTrp 50
Db 735 CCACATCCACATCTACAGTGCAGTGGGCCCCGAGGATGACGGCAAGACAGTGGATGG 794
QY 51 -----ProArgSerAlaSerArgTrpProTrp-----SerAla 61
Db 795 GCACAGCCCGCTGCCGAGGAGGCGCTTCTTCGTGGGTAGCCAGGCGCGGGGGCT 854
QY 62 GlyLeuThrValArgAspArgProGlnLeuGlyGluLeuCysMetGlyArgGly 79
Db 855 GGGCTC-----CATCTTTGTCTGGGCTCGGG 881
RESULT 15
US-08-753-247-20
; Sequence 20, Application US/08753247
; Patent No. 6210929
; GENERAL INFORMATION:
; APPLICANT: SCHLOKAT, Uwe
; APPLICANT: FISCHER, Bernhard
; APPLICANT: FALKNER, Falko-Guenther
; APPLICANT: DORNER, Friedrich
; APPLICANT: EIBL, Johann
; TITLE OF INVENTION: A FUSION PROTEIN COMPRISING A FURIN
; TITLE OF INVENTION: DERIVATIVE OR A DERIVATIVE OF A FURIN ANALOGUE AND A
; TITLE OF INVENTION: HETEROLOGOUS SEQUENCE
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/753,247
; FILING DATE: 22-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AT 1928/95
; FILING DATE: 24-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 40433/149
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1794 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1791
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..1791
US-08-753-247-20
Alignment Scores:
Pred. No.: 48.6 Length: 1794
Score: 85.00 Matches: 30
Percent Similarity: 34.69% Conservative: 4
Best Local Similarity: 30.61% Mismatches: 34
Query Match: 18.64% Indels: 30
DB: 3 Gaps: 6
US-09-972-032-2 (1-79) x US-08-753-247-20 (1-1794)
QY 3 GlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeu 22
Db 615 GGCCAAACACGGTGTCTGTGTAGTGTGGCTTACAGCCCGCATTTGGAGGGGTGG 674
QY 23 TrpThrGly---LeuGlyGluGlyGlnGluGlyGly-----IleGlyProGluGlyGln 39
Db 675 CATGCTGGATGGCGAGGTGCACAGTGCAGTGGAGGACGGTCTGGCTGAACCCCAA 734
QY 40 AlaSerProThrProAspCysAla-----SerArgTrp 50
Db 735 CCACATCCACATCTACAGTGCAGTGGGCCCCGAGGATGACGGCAAGACAGTGGATGG 794
QY 51 -----ProArgSerAlaSerArgTrpProTrp-----SerAla 61
Db 795 GCACAGCCCGCTGCCGAGGAGGCGCTTCTTCGTGGGTAGCCAGGCGCGGGGGCT 854
QY 62 GlyLeuThrValArgAspArgProGlnLeuGlyGluLeuCysMetGlyArgGly 79
Db 855 GGGCTC-----CATCTTTGTCTGGGCTCGGG 881
```

Search completed: July 10, 2004, 18:58:39
Job time : 1527 secs

This Page Blank (uspto)

29: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Run on: July 10, 2004, 16:36:45 ; Search time 2885 Seconds (without alignments)
817.717 Million cell updates/sec

Title: US-09-972-032-2
Perfect score: 456
Sequence: 1 MCGRRPRVSAGCGFADAHWT.....SAGLTVRDPQLGELCMGRG 79

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09972032/runat_06072004_121452_8401/app_query.fasta_1.263
-DB=EST -QFMT=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NOR=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09972032@cgn_1_1_3437 @runat_06072004_121452_8401 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmb:*
5: em_estov:*
6: em_estopl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*

Result No. Score Query Match Length DB ID

Result No.	Score	Query Match	Length	DB ID	Description
1	456	100.0	1467	11	BC039500 Homo sapi
2	421	92.3	1201	13	EX406138 BX406138
3	395	86.6	760	12	BM980506 UI-CF-EN1
4	274	60.1	691	13	BU620427 UI-H-FL1
5	252	55.3	694	13	BU684594 UI-CF-EN1
6	249.5	54.7	1084	12	BM806108 AGENCOURT
7	238	52.2	669	12	BM679577 UI-E-BOO
8	187	41.0	632	13	BU617331 UI-H-DF0
9	187	41.0	633	14	CA313142 UI-CF-FN0
10	183	40.1	623	14	CD742581 UI-H-FT2
11	166	36.4	628	13	BQ447041 UI-H-EUL
12	98.5	21.6	486	12	BM484011 537701 MA
13	96.5	21.2	524	29	CC514259 CH240_359
14	96.5	21.2	1362	13	BQ644396 AGENCOURT
15	95.5	20.9	872	12	BG539339 602567951
16	95.5	20.9	906	14	CD171626 AGENCOURT
17	95.5	20.9	1443	28	BZ575371 meh2_4430
18	95	20.8	563	9	AW047982 UI-M-BH1
19	94.5	20.7	378	13	BU997672 H108K20r
20	94.5	20.7	557	9	AV665399 AV665399
21	94.5	20.7	632	14	CA756603 BR0500020
22	94.5	20.7	664	12	BG860017 1024066H0
23	94.5	20.7	771	14	CB631692 OS1IEB09G
24	94.5	20.7	809	14	CB671698 OSJNE050G
25	94.5	20.7	825	14	CB675995 OSJNE050G
26	94.5	20.7	839	14	CB656383 OSJNE050G
27	94.5	20.7	1343	12	BG284023 602407325
28	94	20.6	282	10	BF364571 PMI-NN108
29	94	20.6	298	29	CE810988 tigr-gss-
30	94	20.6	876	12	BG493231 602541917
31	94	20.6	1114	28	BZ562969 pac62_164
32	93.5	20.5	505	10	B326969 hr8C04.X
33	93.5	20.5	566	9	A1741157 wg19f07.x
34	93.5	20.5	579	13	BQ786882 il51a04.x
35	93.5	20.5	594	28	BH782442 fzmb011f0
36	93.5	20.5	681	29	AY408899 Homo sapi
37	93.5	20.5	686	10	AM968685 EST380761
38	93	20.4	349	29	CE374135 tigr-gss-
39	93	20.4	545	12	B1849019 471654 MA
40	93	20.4	585	9	A1405829 GH25967.5
41	93	20.4	845	10	BF696818 602125353
42	93	20.4	849	29	CNS0329D Tetraodon
43	93	20.4	872	13	BX325065 BX325065
44	92.5	20.3	465	12	B1255478 602977989
45	92.5	20.3	651	28	CC419125 FUHHS48TB

ALIGNMENTS

RESULT 1
BC039500
LOCUS
DEFINITION Homo sapiens, clone IMAGE:5555626, mRNA.
ACCESSION BC039500
VERSION BC039500.1 GI:25058499
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1467)
AUTHORS Strausberg,R.

BC039500 1467 bp mRNA linear HTC 04-MAR-2003

TITLE Direct Submission
JOURNAL Submitted (01-NOV-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
 Email: cgabps-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
 Web site: <http://www-shgc.stanford.edu>
 Contact: (Dickson, Mark) mcd@paxil.stanford.edu
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Place: 88 Row: g Column: 1
 This clone has the following problem: retained intron.

FEATURES

Location/Qualifiers
 1..1467
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:555626"
 /tissue_type="Uterus, leiomyosarcoma"
 /clone_lib="NIH MGC_71"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"

ORIGIN

Alignment Scores:
 Pred. No.: 3,59e-25 Length: 1467
 Score: 456.00 Matches: 79
 Percent Similarity: 100.0% Conservative: 0
 Best Local Similarity: 100.0% Mismatches: 0
 Query Match: 100.0% Indels: 0
 DB: 11 Gaps: 0

US-09-972-032-2 (1-79) x BC039500 (1-1467)

QY 1 MetCysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisThrThr 20
 DB 673 ATGTGTGGGAGACCGCGTCCGTAACGCTGGATGGCTTCGCTGATGCACATTGGACC 732
 QY 21 GlyLeuThrThrGlyLeuGlyGlnGlyGlyLeuGlyGlyLeuGlyGlnAla 40
 DB 733 GGCTCTGACATGGCTAGGGAAGGCGAGGAGGCGGAATTGGCGCCGAGGCGCC 792
 QY 41 SerProThrProAspCysAlaSerArgTTPProArgSerAlaSerArgTTPProThrSer 60
 DB 793 TCGCCGACCCCGACTCGCCCTCCCGTGGCCCGGAGGCGCCCTCCCGTGGCCCTGGAGT 852
 QY 61 AlaGlyLeuThrValArgAspArgProGlnLeuGlyGlnLeuGlyCysMetGlyArgGly 79
 DB 853 GCAGGTCTTACCGTCCGAGATCGTCCGCAACTGGCGAGCTGTGCATGGGGCGTGGC 909

RESULT 2
 BX406138/c
 LOCUS BX406138 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
 DEFINITION BX406138 Homo sapiens FETAL LIVER Homo sapiens cDNA clone
 ACCESSION CS00M010YD19 3-PRIME, mRNA sequence.
 VERSION BX406138
 KEYWORDS BX406138.1 GI:30648317
 EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1201)

AUTHORS

Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 Contact: Genoscope

COMMENT

Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 8395.f For
 more information about this cluster, see
<http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0AM010CB10NP1&cluster=8395.f>. Contact :
 Feng Liang Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com/Invitrogen/Corporation1600>
 Faraday Avenue Genoscope sequence ID : CS0AM010CB10NP1.

FEATURES

source

1..1201
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DM010YD19"
 /tissue_type="FETAL LIVER"
 /dev_stage="fetal"
 /clone_lib="Homo sapiens FETAL LIVER"
 /note="Organ: liver; Vector: pCMVSPORT6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

ORIGIN

Alignment Scores:
 Pred. No.: 1,38e-22 Length: 1201
 Score: 421.00 Matches: 77
 Percent Similarity: 97.47% Conservative: 0
 Best Local Similarity: 97.47% Mismatches: 2
 Query Match: 92.32% Indels: 1
 DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x BX406138 (1-1201)

QY 1 MetCysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisThrThr 20
 DB 713 ATGTGTGGGAGACCGCGTCCGTAACGCTGGATGGTTCGCTGATGCACATTGGACC 654
 QY 21 GlyLeuThrThrGlyLeuGlyGlnGlyGlnGlyGlyLeuGlyGlnAla 40
 DB 653 GGGCTCTGGACTGGGCTAGGGAAGGCGAGGCGGAATTGGCCCGAG-GGCCAGGCC 595
 QY 41 SerProThrProAspCysAlaSerArgTTPProArgSerAlaSerArgTTPProThrSer 60
 DB 594 TCGCCGACCCCGACTCGCCCTCCCGTGGCCCGGAGGCGCCCTCCCGTGGCCCTGGAGT 535
 QY 61 AlaGlyLeuThrValArgAspArgProGlnLeuGlyGlnLeuGlyCysMetGlyArgGly 79
 DB 534 GCAGGTCTTACCGTCCGAGATCGTCCGCAACTGGCGAGCTGTGCATGGGGCGTGGC 478

RESULT 3

BX980506/c
 LOCUS BX980506 760 bp mRNA linear EST 21-FEB-2003
 DEFINITION UI-CF-EN1-add-a-08-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
 UI-CF-EN1-add-a-08-0-UI 3', mRNA sequence.
 ACCESSION BX980506
 VERSION BX980506.1
 KEYWORDS BX980506.1 GI:19602038
 EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 760)
 AUTHORS Ronaldo, M.F., Lennon, G. and Soares, M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL
MEDLINE
PUBMED
COMMENT

Genome Res. 6 (9), 791-806 (1996)
97044477
889548
Contact: McCray, PB
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source

Location/Qualifiers
1. 760
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-EN1-add-a-08-0-UI"
/tissue_type="Primary Lung Cystic Fibrosis Epithelial
Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-EN1"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-CF-EN1 is a normalized cDNA library containing the
following tissue(s): Primary Lung Cystic Fibrosis
Epithelial Cells. The library was constructed according to
Bonald, Lennon and Soares, Genome Research, 6:791-806,
1996. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was ligated to an EcoR I adaptor, digested with Not
I, and cloned directionally into pT73-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CTGCTCAGGT.
TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS
6hr to LPS 24h
TAG_LIB=UI-CF-EN1
TAG_SEQ=CTGCTCAGGT"

ORIGIN

Alignment Scores:
Pred. No.: 7.63e-21 Length: 760
Score: 395.00 Matches: 73
Percent Similarity: 97.33% Conservativeness: 0
Best Local Similarity: 97.33% Mismatches: 2
Query Match: 86.62% Indels: 1
DB: 12 Gaps: 0

US-09-972-032-2 (1-79) x BM980506 (1-760)

Qy 5 ProArgValSerAlaGlyCysGlyPheAlaAspAlaHisThrGlyLeuTrpThr 24
Db 758 CCGCGTCCGTAAGCGTGGATGTC-TTCGCTGATGCACATTGCACCGGCTCTGGACT 700
Qy 25 GlyLeuGlyGluGlyGlnGlyGlyLeuGlyGlnAlaSerProThrPro 44
Db 699 GGGCTAGGGAGAGGAGAGGCGGAATGGGCCCCGAGGGCGAGGCTCGCCAGCCCC 640
Qy 45 AspCysAlaSerArgTrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThr 64
Db 639 GANTGGCCCTCCCGTGGCCCGCAGCGCTCCCGTGGCCCTCGAGTGCAGGTCTTACC 580
Qy 65 ValArgAspArgProGlnLeuGlyGluLeuCysMetGlyArgGly 79

Db 579 GTCCGAGATCGTCGCAACTGGCGAGCTGTCATGGCGGTGGC 535

RESULT 4
LOCUS

BU620427/c
DEFINITION
UI-H-FLI-bfw-n-19-0-UI-s1 NCI_CGAP_FLI Homo sapiens cDNA clone
UI-H-FLI-bfw-n-19-0-UI 3', mRNA sequence.
ACCESSION
BU620427
VERSION
BU620427.1 GI:23286642
KEYWORDS
EST.
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
1 (bases 1 to 691)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: James Martin
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bentos-soares@uiowa.edu
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
source

Location/Qualifiers
1. 691
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-FLI-bfw-n-19-0-UI"
/tissue_type="Cell lines"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP_FLI"
/note="Organ: Chondrosarcoma; Vector: pT73-Pac
(Pharmacia) with a modified polylinker; Site_1: EcoR I;
Site_2: Not I; NCI_CGAP_FLI is a normalized cDNA library
derived from a pool of mRNA obtained from 4 cell lines
from grade III chondrosarcoma tissues. The library was
constructed according to Bonaldo, Lennon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT73-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
GAGTCCGCTG. The cell lines were provided by Dr. James
Martin from the University of Iowa.
TAG TISSUE=Human Chondrosarcoma Grade 3 cell line mix
TAG_LIB=UI-H-FLI
TAG_SEQ=GAGTCCGCTG"

ORIGIN

Alignment Scores:
Pred. No.: 1.45e-11 Length: 691
Score: 274.00 Matches: 52
Percent Similarity: 98.11% Conservativeness: 0
Best Local Similarity: 98.11% Mismatches: 1
Query Match: 60.09% Indels: 1
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x BU620427 (1-691)

Qy 27 GlyGluGlyGlnGluGlyGlyLeuGlyProGluGlyGlnAlaSerProThrProAspCys 46

```

Db      689  GGGAGAGGCGAGGCGGGAAT-GGGCCCCAGAGCGGCGCTCGCCGACCCCGACTGC 631
Qy      47  AlaSerArgTrpProAlaSerAlaSerArgTrpProTrpSerAlaGlyLeuThrValArg 66
Db      630  GCCTCCCGGTGGCCCGCAGCGCTCCCGGTGGCCCTGGAGTGGAGTCTTACCGTCCGA 571
Qy      67  AspArgProGlnLeuGlyGluLeuCysMetGlyArgGly 79
Db      570  GATCGTCGCACTGGCGGAGCTGTGATGGCGGTGGC 532

RESULT 5
BU684594/c
LOCUS      BU684594
DEFINITION UI-CF-EN1-act-a-22-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
VERSION      BU684594
KEYWORDS      BU684594.1 GI:23537704
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL      97044477
MEDLINE      8889548
PUBMED
COMMENT      Contact: McCray, PB
                University of Iowa
                2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
                Tel: 319 356 4866
                Fax: 319 356 7171
                Email: paul-mccray@uiowa.edu
                Tissue Procurement: Dr. M. J. Welsh, University of Iowa
                cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
                cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
                DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
                Clone Distribution: Researchers may obtain clones from Research
                Genetics (www.resgen.com) or from Open Biosystems
                (www.openbiosystems.com).
                Seq primer: M13 FORWARD
                POLYA=Yes.

FEATURES             Location/Qualifiers
     source           1..694
     organism="Homo sapiens"
     mol_type="mRNA"
     db_xref="taxon:9606"
     clone="UI-CF-EN1-act-a-22-0-UI"
     tissue_type="Primary Lung Cystic Fibrosis Epithelial
     Cells"
     dev_stage="Adult"
     lab_host="DH10B (Life Technologies) (71 phage resistant)"
     clone_lib="UI-CF-EN1"
     notes="Organ: Lung; Vector: p7T73-Pac (Pharmacia) with a
     modified polylinker; Site 1: EcoR I; Site 2: Not I;
     UI-CF-EN1 is a normalized cDNA library containing the
     following tissue(s): Primary Lung Cystic Fibrosis
     Epithelial Cells. The library was constructed according to
     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
     1996. First strand cDNA synthesis was primed with an
     oligo-dT primer containing a Not I site. Double stranded
     cDNA was ligated to an EcoR I adaptor, digested with Not
     I, and cloned directionally into p7T73-Pac vector. The
     oligonucleotide used to prime the synthesis of
     first-strand cDNA contains a library tag sequence that is
     located between the Not I site and the (dT)18 tail. The
     sequence tag for this library is CTGCTCAGGT.
     TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS
     6hr to LPS 24h
     DB: LIB=UI-CF-EN1

```

TAG_SEQ=CTGCTCAGGT"

ORIGIN

```

Alignment Scores:
Pred. No.:      7,248-10      Length:      694
Score:          252.00        Matches:      48
Percent Similarity: 96.00%    Conservative: 0
Best Local Similarity: 96.00% Mismatches:      2
Query Match:     55.26%      Indels:       1
DB:              13          Gaps:         0

```

US-09-972-032-2 (1-79) x BU684594 (1-694)

```

Qy      30  GlnGluGlyGlyLeGlyProGluGlyGlnAlaSerProThrProAspCysAlaSerArg 49
Db      694  CAGGAGGGCGGAAT-GGGCCCCAGAGCGGCGCTCGCCGACCCCGANTGGCGCTCCCGG 636
Qy      50  TrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThrValArgAspArgPro 69
Db      635  TGGCCCCCGAGCGCTCCCGGTGGCCCTGGAGTGGAGTCTTACCGTCCGAGATCGTCCG 576
Qy      70  GlnLeuGlyGluLeuCysMetGlyArgGly 79
Db      575  CAACTGGCGGAGCTGTGATGGCGGTGGC 546

```

RESULT 6

BM806108

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

US-09-972-032-2 (1-79) x BM806108 (1-1084)

Qy 1 MetCysGlyArgProArgArgValSerAlaGlyCysGlyPheAlaaspAlaHisTrpThr 20
 Db 672 ATGTGGGAGACCGCTCGCTAGCGCTGGATGGCTTCCTGATGACATTGGACC 731
 Qy 21 GlyLeuTrpThrGlyLeuGlyGlu-GlyGlnGluGlyGlyLe----GlyProGluGlyG 39
 Db 732 GGGCTCGGACTGGGCTAGGGAAAGGCCAGAGGGCCGGATTGGGGGGCCCGAGGGCCA 791
 Qy 39 nAlaSerProThrProAspCysAla-SerArgTrp-ProArgSerAlaSerArgTrp--P 58
 Db 792 GGCTTCGCCAGCCCCCCTCGGCGCTCCCGTGGGGCCGACAGGGCGCTCCCGTGGGCC 851
 Qy 58 roTrpSerAla---GlyLeuThrValArgAspArgProGlnLeuGlyGluLeuCys---M 76
 Db 852 CCTGGGAGGCAAGCGCTTACCGCTCCCAATCGTCCCAACCTGGGCGCAATTGGGCT 911
 Qy 76 etGlyArgGly 79
 Db 912 GGGGGCGGGGC 922

RESULT 7
 BM679577/c
 LOCUS BM679577 669 bp mRNA linear EST 27-FEB-2002
 DEFINITION UI-E-E00-aia-1-05-0-UI.sl UI-E-E00 Homo sapiens cDNA clone
 UI-E-E00-aia-1-05-0-UI 3', mRNA sequence.
 ACCESSION BM679577
 VERSION BM679577.1 GI:18989473
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Bernaldo,M.F., Lennon,G. and Soares,M.B.
 1 (bases 1 to 669)
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 PUBMED 8889548
 COMMENT Contact: Soares, MB
 Coordinated Laboratory for Computational Genomics
 University of Iowa
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: bento-soares@uiowa.edu
 Tissue Procurement: Dr. Gregg Hageman
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com).
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES
 source Location/Qualifiers
 1..669
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-E-E00-aia-1-05-0-UI"
 /tissue_type="fetal eye"
 /dev stage="fetal"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_lib="UI-E-E00"
 /notes="Organ: eye; Vector: p7T3-Pac (Pharmacia) with a modified polylinker; Site:1: EcoR I; Site 2: Not I; UI-E-E00 is a cDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bernaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed

with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dri)18 tail. The sequence tag for this library is CCGGTATACC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).
 TAG TISSUE=human fetal eye
 TAG LIB=UI-E-E00
 TAG_SEQ=CGGTATACC"

ORIGIN
 Alignment Scores:
 Pred. No.: 8,29e-09 Length: 669
 Score: 238.00 Matches: 42
 Percent Similarity: 97.67% Conservative: 0
 Best Local Similarity: 97.67% Mismatches: 1
 Query Match: 52.19% Indels: 0
 DB: 12 Gaps: 0

US-09-972-032-2 (1-79) x BM679577 (1-669)

Qy 37 GluGlyGlnAlaSerProThrProAspCysAlaSerArgTrpProArgSerAlaSerArg 56
 Db 669 GAGGCGCCAGGCTCGCCAGCCCGGANTGCGCTCCCGGCGCCGAGCGCCCTCCCGG 610
 Qy 57 TrpProTrpSerAlaGlyLeuThrValArgAspArgProGlnLeuGlyGluLeuCysMet 76
 Db 609 TGGCCCTGGAGTGCAGGCTTACCGTCCGAGATCGTCCGCACTGGGCGAGCTGTGCATG 550
 Qy 77 GlyArgGly 79
 Db 549 GGGCGTGGC 541

RESULT 8
 BU617331/c
 LOCUS BU617331 632 bp mRNA linear EST 23-SEP-2002
 DEFINITION UI-H-DF0-bep-n-09-0-UI.sl NCI-CGAP_DFO Homo sapiens cDNA clone
 UI-H-DF0-bep-n-09-0-UI 3', mRNA sequence.
 ACCESSION BU617331
 VERSION BU617331.1 GI:23283539
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 632)
 REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapsb-remail.nih.gov
 Tissue Procurement: Dr. Jose Mercuende
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES
 source Location/Qualifiers
 1..632
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-H-DF0-bep-n-09-0-UI"
 /tissue_type="Subchondral Bone"
 /dev stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI-CGAP_DFO"

/note="Organ: Bone; Vector: p773-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; NCI CGAP DFO is a cDNA library containing the following tissue(s): Subchondral Bone. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is GTTAAGCGTC.
TAG_TISSUE=subchondral bone
TAG_LIB=UI-H-DFO
TAG_SEQ=GTTAAGCGTC"

ORIGIN

Alignment Scores:
Pred. No.: 6.6e-05 Length: 632
Score: 187.00 Matches: 33
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 41.01% Indels: 0
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x BU617331 (1-632)

Qy 47 AlaSerArgTrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThrValArg 66
Db 630 GCCTCCCGGTGGCCCGCAGCGCTCCCGTGGAGTGCAGGTCTTACCGTCCGA 571
Qy 67 AspArgProGlnLeuGlyGluLeuCysMetGlyArgGly 79
Db 570 GATCGTCCGCACTGGCGGAGCTGTGCATGGGCGGTGGC 532

RESULT 9

CA313142/c
LOCUS CA313142 633 bp mRNA linear EST 04-NOV-2002
DEFINITION UI-CF-FNO-aex-n-23-0-UI.s1 UI-CF-FNO Homo sapiens cDNA clone
UI-CF-FNO-aex-n-23-0-UI 3', mRNA sequence.

ACCESSION CA313142.1 GI:24531240
VERSION CA313142.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 633)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)
AUTHORS 97044477
TITLE 889548
PUBMED

COMMENT Contact: McCray, PB
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu

Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com) or from Open Biosystems (www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=Yes.
Location/Qualifiers
1. .633

FEATURES

source

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="UI-CF-FNO-aex-n-23-0-UI"
/tissue_type="Human Lung Epithelial cells"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-FNO"
/note="Organ: Lung; Vector: p773-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-CF-FNO is a subtracted cDNA library derived from two normalized Human lung epithelial cell libraries (BN1 and DUL) The library was subtracted according to Bonaldo, Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. For additional information, contact: bentco-soares@uiowa.edu
TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS
6hr to LPS 24h
TAG_LIB=UI-CF-FNO
TAG_SEQ=CTGCTCAGGT"

ORIGIN

Alignment Scores:
Pred. No.: 6.61e-05 Length: 633
Score: 187.00 Matches: 33
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 41.01% Indels: 0
DB: 14 Gaps: 0

US-09-972-032-2 (1-79) x CA313142 (1-633)

Qy 47 AlaSerArgTrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThrValArg 66
Db 633 GCCTCCCGGTGGCCCGCAGCGCTCCCGTGGAGTGCAGGTCTTACCGTCCGA 574
Qy 67 AspArgProGlnLeuGlyGluLeuCysMetGlyArgGly 79
Db 573 GATCGTCCGCACTGGCGGAGCTGTGCATGGGCGGTGGC 535

RESULT 10

CD742581/c

LOCUS CD742581 623 bp mRNA linear EST 26-JUN-2003
DEFINITION UI-H-FT2-bj1-i-10-0-UI.s1 NCI CGAP_F22 Homo sapiens cDNA clone
UI-H-FT2-bj1-i-10-0-UI 3', mRNA sequence.

ACCESSION CD742581.1 GI:32293431
VERSION CD742581
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 623)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: ccgaps-remail.nih.gov
Tissue Procurement: Dr. Gary W. Hunninghake, U of I
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Distribution information can be found at
<http://genome.uiowa.edu/distribution/cgap.html>
Seq primer: M13 FORWARD
POLYA=Yes.
Location/Qualifiers
1. .623

FEATURES

source

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clones="UI-H-FT2-bj1-i-10-0-UI"

/tissue_type="Aveolar Macrophage"

```

/dev stage="Adult"
/lab host="DH10B (Life Technologies)"
/clone lib="NCI_CGAP_F72"
/note="Organ: Lung; Vector: p773-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
NCI_CGAP_F72 is a subtracted cDNA library constructed from
a pool of 81 RNA samples from Alveolar Macrophages
challenged with different treatments. The library was
subtracted according to Bonaldo, Lennon and Soares, Genome
Research, 6:791-806, 1996. The tissue was provided by Dr.
Gary W. Hunninghake of the University of Iowa.
TAG_TISSUE=Human Lung Alveolar Macrophage
TAG_LIB=UI-H-FT2
TAG_SEQ=GGCCATGGCG"

```

ORIGIN

```

Alignment Scores:
Pred. No.: 0.000132 Length: 623
Score: 183.00 Matches: 32
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 40.13% Indels: 0
DB: 14 Gaps: 0

```

US-09-972-032-2 (1-79) x CD742581 (1-623)

```

Qy 48 SerArgTrpProArgSerAlaSerArgTrpProTrpSerAlaGlyLeuThrValArgAsp 67
Db 622 TCCCGGTGGCCCGCAGCGCTCCCGTGGCCCTCGAGTGCAGTCTTACCGTCCGAGAT 563
Qy 68 ArgProGlnLeuGlyGluLeuCysMetGlyArgGly 79
Db 562 CGTCCGCAACTGGCGGAGCTGTGCATGGGCGGTGGC 527

```

RESULT 11

```

BQ447041/c
LOCUS BQ447041 628 bp mRNA linear EST 29-MAY-2002
DEFINITION UI-H-EUI-bac-p-06-0-UI.s1 NCI_CGAP_Ctl Homo sapiens cDNA clone
ACCESSION BQ447041
VERSION BQ447041.1 GI:21250153
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 628)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapsb@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
Seq primer: M13 FORWARD
POLVA=Yes.
Location/Qualifiers
1..628
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-H-EUI-bac-p-06-0-UI"
/tissue type="Osteoarthritic Cartilage"
/dev stage="Adult"
/lab host="DH10B (Life Technologies)"
/clone lib="NCI_CGAP_Ctl"
/note="Organ: Knee; Vector: p773-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;

```

FEATURES

source

NCI_CGAP_Ctl is a normalized cDNA library containing the following tissue(s): Osteoarthritic Cartilage The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is TGATCAGCGT.

TAG_TISSUE=osteoarthritic cartilage
TAG_LIB=UI-H-EUI
TAG_SEQ=TGATCAGCGT"

ORIGIN

```

Alignment Scores:
Pred. No.: 0.00273 Length: 628
Score: 166.00 Matches: 31
Percent Similarity: 96.88% Conservatives: 0
Best Local Similarity: 96.88% Mismatches: 0
Query Match: 36.40% Indels: 1
DB: 13 Gaps: 0

```

US-09-972-032-2 (1-79) x BQ447041 (1-628)

```

Qy 49 ArgTrpProArgSerAlaSerArg-TrpProTrpSerAlaGlyLeuThrValArgAsp 68
Db 626 CGTGGCCCGCAGCGCTCCCGTGGCCCTCGAGTGCAGTCTTACCGTCCGAGATCG 567
Qy 68 gProGlnLeuGlyGluLeuCysMetGlyArgGly 79
Db 566 TCCGCAACTGGCGGAGCTGTGCATGGGCGGTGGC 533

```

RESULT 12

```

BQ484011/c
LOCUS BQ484011 486 bp mRNA linear EST 05-FEB-2002
DEFINITION 537701 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BQ484011
VERSION BQ484011.1 GI:18534339
KEYWORDS EST.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 486)
AUTHORS Fahrenkrug, S.C., Smith, T.P.L., Freking, B.A., Cho, J., White, J.,
Vallet, J., Wise, T., Rohrer, G.A., Perte, G., Sultana, R.,
Quackenbush, J., and Keele, J.W.
Porcine gene discovery by normalized cDNA-library sequencing and
EST cluster assembly
Mamm. Genome 13 (8), 475-478 (2002)
22213789
12226715
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCATGACCAT
BACKWARD: GTTTCAGTCACGACG
Plate: 2 row: K column: 16
Seq primer: ATTTAGTGACACTATAG.
Location/Qualifiers
1..486
/organism="Sus scrofa"
/mol_type="mRNA"

```

```

/db_xref="taxon:9823"
/tissue_type="pooled"
/lab_host="DH103"
/clone_lib="MARC 2PG"
/notes="Vector: pCMV SPOR6; Site 1: NotI; Site 2: SalI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."

ORIGIN

Alignment Scores:
Pred. No.:      312      Length:      486
Score:          98.50      Matches:      29
Percent Similarity: 44.74%      Conservative: 5
Best Local Similarity: 38.16%      Mismatches: 22
Query Match:      21.60%      Indels:      20
DB:               12         Gaps:       5

US-09-972-032-2 (1-79) x BM484011 (1-486)

QY      2 CysGlyArgProArgValSerAlaGlyCys-GlyPheAlaAspAlaHisTrpThrGl 21
Ddb     470 TCGCGT---CCGAGATGCTCTTCAGATGCCGATACCACTGCCACCAAGTCGA 414

QY      21 yLeuTrpThr-----GlyLeuGlyGluGlyGlnGluGlyGlyIleGly---ProGl 37
Ddb     413 TCTTTGGACCGTCACCCCTGGCGTGGTCAGGGTCCACGGGTCTTCGGGCTCTCTCG 354

QY      37 uGlyGlnAlaSerProThrProAspCysAlaSerArgTrpProArgSerAlaSerArg-- 56
Ddb     353 GGGGCAGCGGCGCC-----CGCGTCTGCTGAGAGAAAT 324

QY      57 -----TrpProTrpSerAlaGlyLeuThrValArgAspArg 68
Ddb     323 TGGTCTTGGGTGGCGCGTGGCTTTTGGGTGGTATGGCAGGACAGA 278

CCS14259      524 bp      DNA      linear      GSS 17-JUN-2003
CH240_359M23_T7 CHORI-240 Bos taurus genomic clone CH240_359M23,
genomic survey sequence.
CCS14259      GI:31832547
CCS14259.1
GSS.
Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (Bases 1 to 524)
Holt, R., Stott, J., Yang, G., Barber, S., Smailus, D., Prabhu, A.-L.,
Tsai, M., Cloutier, A., Lee, D., Girm, N., Olson, T., Mayo, M.,
Butterfield, Y., Kirkpatrick, R., Liu, J., Guin, R., Chan, A., Chiu, R.,
Mathewson, C., Wye, N., Masson, A., Brown-John, M., Jones, S.,
Schein, J., Warra, M., de Jong, P., McWilliam, S., Barris, W.,
Dairymple, B.P. and Tellam, R.
Bovine BAC End Sequences from Library CHORI-240, PLATES 294 to 398
Unpublished (2003)
Other GSSs: CH240_359M23.TARBAC13P2
Contact: Rob Holt
Sequencing
The British Columbia Cancer Agency Genome Science Centre
600 W. 10th Ave, Vancouver, British Columbia, Canada V5Z 4S6
Tel: 604-877-6085
Fax: 604-877-6276
Email: rholt@ccgc.ca
Clones are derived from the bovine BAC library CHORI-240
(http://www.chori.org/bacpac/bovine240.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering/information.html). This work
was undertaken as part of the International Bovine BAC Mapping
Consortium (IBBMC) by CSIRO Livestock Industries, Australia and the
British Columbia Genome Sciences Centre, Canada.

```

http://image.llnl.gov
Plate: LLCM1513 row: e column: 14
High quality sequence stop: 591.
Location/Qualifiers
1..872

FEATURES

source

1. 872
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:46246"
/lab_host="PH10B (T1 phage-resistant)"
/clone_lib="NIH_MGC_77"
/notes="Organ: lung; Vector: pDNR-LIB (SfiI (ggcgctctggcg); Site 2: SfiI (ggcgctctggcg) were used in cloning as for 3' adaptors were used in cloning as for sequence: 5'-CAGCGCCATTATGCC-3' and 5'-ATTCTAGCGCGCGCGGCACATG-dt(30)B12C, or G and N = A, C, G, or T). Averted 12/15 colonies
kb (range 0.5-4.0 kb). 12/15 colonies by PCR. This library was enriched for and was constructed by Clontech Laboratory and was constructed by Clontech Laboratory. CA). Note: this is a NIH MGC Library."

ORIGIN

Alignment Scores:	
Pred. No.:	1.15e+03
Score:	95.50
Percent Similarity:	36.79%
Best Local Similarity:	30.19%
Query Match:	20.94%
DB:	12
Gaps:	5
Indels:	37
Mismatches:	30
Conservative:	7
Matches:	32
Length:	872

US-09-972-032-2 (1-79) X BG539339 (1-872)

2	CysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGly	21
	::::	:::
207	TGTGCCAGACGCTGGAGATGGCGCAGCGCAGCGTGTGGGGTGTGGAGTCTCTGTGTGC	266
	::::	:::
22	LeuTrpThrGlyLeuGlyGlnGluGlyGlyGlyProGlu	37
	::::	:::
267	TGTTGGACAGACCCGGGAGGCCCA	320
	::::	:::
38	-----GlyGlnAlaSerProThrPro	44
	::::	:::
321	ACGCACATGACTGTGAGGCCCTGGGCTGGGGTCCGAGGCAGCCGCTCCAGGAGTCCC	380
	::::	:::
45	Asp-----CysAlaSerArg-----TrpProArg	52
	::::	:::
381	GAGCACCGTCAAACTGCAGTTCGATATGTCGGCCCTGCACCTGGTGGCCACGGCGG	440
	::::	:::

RESULT 15
BG539339 872 bp mRNA linear EST 03-APR-2001
602567951F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4692469 5',
mRNA sequence.

Accession: BG539339
Accession: BG539339.1
Version: GI:13531572

KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 872)

REFERENCE
1 (bases 1 to 872)
AUTHORS
NTH-MGC <http://mgc.nci.nih.gov/>.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)

CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-i@mail.nih.gov
Tissue Procurement: CLONTECH Laboratories, Inc.
cDNA Library Preparation: CLONTECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E Consortium/LLNL at:

found through the I.M.A.G.E. Consortium/LLNL at:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 10, 2004, 15:31:05 ; Search time 2821 Seconds

(without alignments)
1213.789 Million cell updates/sec

Title: US-09-972-032-2

Perfect score: 456

Sequence: 1 MCGRRRVSGAGCFADAHWT.....SAGLTVRDPQLGELCMGRG 79

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgm2 1/USPTO.spool/US09972032/runat_06072004_121451_8389/app.query.fasta_1.263
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NOR=ext -HEA=SIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09972032 @CGN 1 1 3731 @runat_06072004_121451_8389 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hgt:*
3: gb_in:*
4: gb_em:*
5: gb_ov:*
6: gb_pat:*
7: gb_pl:*
8: gb_pr:*
9: gb_ro:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_hgt_hum:*
31: em_hgt_inv:*
32: em_hgt_other:*
33: em_hgt_mus:*
34: em_hgt_pln:*
35: em_hgt_rod:*
36: em_hgt_mam:*
37: em_hgt_vrt:*
38: em_by:*
39: em_hgt_hum:*
40: em_hgt_mus:*
41: em_hgt_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	456	100.0	182230	9	AC135050	Homo sapi
2	456	100.0	195476	2	AC135044	Homo sapi
C 3	105	23.0	3628	8	AK110169	Oryza sat
C 4	103	22.6	163035	2	AC006405	Homo sapi
5	100	21.9	900	9	SS3698	tal-1-exons
6	99.5	21.8	136098	9	AC006970	Homo sapi
C 7	98.5	21.6	2163	5	AF164165	Certhidea
C 8	98.5	21.6	153675	2	AC092012	Felis cat
C 9	98	21.5	185997	2	AC134385	Papio anu
10	98	21.5	191292	2	AC138018	Homo sapi
C 11	98	21.5	303550	1	SC093913	Streptomy
12	97.5	21.4	1476	6	BD179739	Highly th
13	97.5	21.4	45858	9	AC006123	Homo sapi
C 14	97.5	21.4	283100	1	SC093911	Streptomy
C 15	97	21.3	193012	2	AC113007	Mus muscu
16	97	21.3	300800	1	SC093912	Streptomy
17	96	21.1	161090	9	AC117415	Homo sapi
18	96	21.1	193265	2	AC138705	Homo sapi
19	96	21.1	308147	1	AE016915	Chromobac
C 20	95.5	20.9	1325	8	AK106138	Oryza sat
21	95.5	20.9	72907	2	AC016010	Homo sapi
22	95.5	20.9	143961	8	AC018929	Oryza sat
23	95.5	20.9	300029	8	AE017122	Oryza sat
24	95	20.8	44491	9	AC073308	Homo sapi
25	95	20.8	88733	2	AC132855	Homo sapi
26	95	20.8	151495	2	AC068240	Homo sapi
C 27	95	20.8	181671	9	AC007040	Homo sapi
C 28	95	20.8	345997	1	BX569692	Synechoco
C 29	95	20.8	349901	6	AX647201	Sequence
C 30	94.5	20.7	809	4	BTGST	Bovine GST
C 31	94.5	20.7	1722	8	AK060525	Oryza sat
32	94.5	20.7	24081	6	AX598605	Sequence
33	94.5	20.7	52101	6	AX598593	Sequence
34	94.5	20.7	131424	8	AF004815	Oryza sat
C 35	94.5	20.7	140241	9	AC079355	Homo sapi
C 36	94.5	20.7	188121	2	AC093394	Bos tauru
C 37	94.5	20.7	227595	2	AC093396	Bos tauru
38	94.5	20.7	230138	14	AF232689	Rat cytom
39	93.5	20.5	606	6	AX119213	Sequence
40	93.5	20.5	943	6	AX574584	Sequence
41	93.5	20.5	943	9	AX358178	Homo sapi
42	93.5	20.5	1641	9	BC004932	Homo sapi
43	93.5	20.5	1870	9	AK024551	Homo sapi
44	93.5	20.5	3161	9	AK126094	Homo sapi
45	93	20.4	45886	2	AC017393	Drosophil

ALIGNMENTS

RESULT 1

```

AC135050      182230 bp      DNA      linear      PRI 27-FEB-2003
LOCUS      Homo sapiens chromosome 16 clone RP11-196G11, complete sequence.
ACCESSION      AC135050
VERSION      AC135050.3 GI:28570306
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 182230)
AUTHORS      DOE Joint Genome Institute, Stanford Human Genome Center and Los
              Alamos National Laboratory.
TITLE      Direct Submission
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 182230)
AUTHORS      DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL      Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE      3 (bases 1 to 182230)
AUTHORS      DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL      Submitted (22-OCT-2002) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE      4 (bases 1 to 182230)
AUTHORS      DOE Joint Genome Institute, Stanford Human Genome Center and Los
              Alamos National Laboratory.
TITLE      Direct Submission
JOURNAL      Submitted (27-FEB-2003) DOE Joint Genome Institute, 2800 Mitchell
              Drive, Walnut Creek, CA 94598, USA
COMMENT      On Feb 27, 2003 this sequence version replaced gi:24211094.
              Draft Sequence Produced by DOE Joint Genome Institute
              www.jgi.doe.gov
              Finishing Completed at Stanford Human Genome Center and Los Alamos
              National Laboratory
              www.shgc.stanford.edu
              Quality: Phrap Quality >=40 99.7% of Sequence;
              Estimated Total Number of Errors is 0.3.
FEATURES             Location/Qualifiers
     source            1..182230
                       /organism="Homo sapiens"
                       /mol_type="genomic DNA"
                       /db_xref="taxon:9606"
                       /chromosome="16"
                       /clone="RP11-196G11"
ORIGIN
Alignment Scores:
Pred. No.:      5.23e-21      Length:      182230
Score:          456.00      Matches:      79
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     100.00%      Indels:      0
DB:              9      Gaps:      0
US-09-972-032-2 (1-79) x AC135050 (1-182230)
Qy      1 MetCysGlyArgProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThr 20
      |||
Db      98089 ATGTGTGGAGACCGCGTCGTAAGCGTGGATGTGGCTTCGTCATGCACATTGCACC 98148
      |||
Qy      21 GlyLeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlyGlyGlyGlyGlyGlnAla 40
      |||
Db      98149 GGGCTCTGGACTGGCTAGGGGAAGGGCAGAGGGGGGGAATTGGGGCCCGAGGCCAGGCC 98208
      |||
Qy      41 SerProThrProAspCysAlaSerArgTrpProArgSerAlaSerArgTrpProTrpSer 60
      |||
Db      98209 TCGCCGACCCCGGACTCGGCTCCCGGTGGCCCCCGCAGCGCCCTCCCGGTGGCCCTGAGT 98268
      |||
Qy      61 AlaGlyLeuThrValArgAspArgProGlnLeuGlyGlnLeuGlyCysMetGlyArgGly 79
      |||
Db      98269 GCAGGTCTTACCGTCCGAGATCGTCCGCAACTGGCGAGCTGTGCATGGGGGTGGC 98325
      |||

```

```

RESULT 2
AC135044      195476 bp      DNA      linear      HTG 05-OCT-2002
LOCUS      Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT
DEFINITION      SEQUENCE, 22 unordered pieces.
ACCESSION      AC135044
VERSION      AC135044.1 GI:23505535
KEYWORDS      HTG; HTGS_PHRASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 195476)
AUTHORS      DOE Joint Genome Institute.
TITLE      Sequencing of Human Chromosome 16
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 195476)
AUTHORS      DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL      Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT      -----Genome Center
              Center: Joint Genome Institute
              Center Code: JGI
              Web site: http://www.jgi.doe.gov
              -----
              Project Information
              Center Project Name: 809609
              Center Clone Name: CTD-EL_2551B20
              -----
              Summary Statistics
              Consensus quality: 171229 bases at least Q40
              Consensus quality: 182638 bases at least Q30
              Consensus quality: 188095 bases at least Q20
              Estimated insert size: 170000; agarose-fp estimation
              Estimated insert size: 193376; sum-of-contigs estimation
              Quality coverage: 7.63 in Q20 bases; agarose-fp estimation
              Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation.
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 22 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence.
              * as soon as it is available and the accession number will
              * be preserved.
              *
              1      1202: contig of 1202 bp in length
              *      1203      1302: gap of unknown length
              *      1303      2468: contig of 1166 bp in length
              *      2469      2568: gap of unknown length
              *      2569      4077: contig of 1509 bp in length
              *      4078      4177: gap of unknown length
              *      4178      5772: contig of 1595 bp in length
              *      5773      5872: gap of unknown length
              *      5873      7722: contig of 1850 bp in length
              *      7723      7822: gap of unknown length
              *      7823      8958: contig of 1136 bp in length
              *      8959      9058: gap of unknown length
              *      9059      11051: contig of 1993 bp in length
              *      11052      11151: gap of unknown length
              *      11152      13641: contig of 2490 bp in length
              *      13642      13741: gap of unknown length
              *      13742      1738: contig of 3997 bp in length
              *      17339      17838: gap of unknown length
              *      17839      22030: contig of 4192 bp in length
              *      22031      22130: gap of unknown length
              *      22131      27002: contig of 4872 bp in length
              *      27003      27102: gap of unknown length
              *      27103      32819: contig of 5717 bp in length
              *      32820      32919: gap of unknown length
              *      32920      39003: contig of 6084 bp in length
              *      39004      39103: gap of unknown length

```



```

source      1. 3628
            /organism="Oryza sativa (japonica cultivar-group)"
            /mol_type="mRNA"
            /cultivar="Nipponbare"
            /db_xref="taxon:39947"
            /clone="002-161-G03"

ORIGIN
Alignment Scores:
Pred. No.: 170 Length: 3628
Score: 105.00 Matches: 38
Percent Similarity: 39.09% Conservative: 5
Best Local Similarity: 34.55% Mismatches: 25
Query Match: 23.03% Indels: 42
DB: 8 Gaps: 9

US-09-972-032-2 (1-79) x AK110169 (1-3628)
Qy 2 CysGlyArg-----ProArgValSerAla----- 10
Db 998 TCGGTCGTACTGATCGCGGTCTGGACGACCTCGGCACCTCGTGGCGCTCTGTAGG 939
Qy 11 GlyCysGlyPheAlaaspAlaH:stTrrThrGlyLeuTrrThrGlyLeuGly----- 27
Db 938 GGACCTTGA-----TGGACTGTTAGTTGGACGATCGAGGTTGGTGACCA 894
Qy 28 GluGly---GlnGluGlyGlyIleGlyProGluGlyGlnAlaSerProThrProaspCys 46
Db 893 GAAGTTGGCGAGGAGCGCTGGAGCTGGTGGACGACATATAGCGCGCTTGTGT 834
Qy 47 AlaSerArgTrrProArgSerAlaSerArgTrrProTrr----- 59
Db 833 CGAGTTCGTTGG---CGATGACGAGACCGTGGCATGGGTTGTAGTGTCTGGCGCCG 777
Qy 60 -----SerAlaGlyLeuThr-----ValArgAspArgPro 69
Db 776 AGTGAAGGCGAGTGGAGCGATCGATGAGTGTGCGGTCTTGGATCCGGAGCGG---- 720
Qy 70 GlnLeuGlyGluLeuCysMetGlyArgGly 79
Db 719 -----CGCACATGTCGAGCAGCAGGT 699

RESULT 4
AC006405/c
LOCUS
DEFINITION Homo sapiens chromosome 17 clone hRPK.107 N.19 map 17, ***
SEQUENCING IN PROGRESS ***, 12 unordered pieces.
AC006405
AC006405.1 GI:4165361
HTG; HTGS_PHASE1.
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 163035)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone hRPK.107_N_19
Unpublished
2 (bases 1 to 163035)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Baria,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Collangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArelano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaton,D., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,
Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Testaye,S., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,

```

```

Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
Direct Submission
Submitted (15-JAN-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 21, 1999 this sequence version replaced gi:4159875.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 12 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 8815: contig of 8815 bp in length
* gap of unknown length
* 8816 12285: contig of 3470 bp in length
* gap of unknown length
* 12286 40956: contig of 28671 bp in length
* gap of unknown length
* 40957 64393: contig of 23437 bp in length
* gap of unknown length
* 64394 78975: contig of 14582 bp in length
* gap of unknown length
* 78976 96637: contig of 17662 bp in length
* gap of unknown length
* 96638 105967: contig of 9330 bp in length
* gap of unknown length
* 105968 145832: contig of 39865 bp in length
* gap of unknown length
* 145833 147041: contig of 1209 bp in length
* gap of unknown length
* 147042 150498: contig of 3457 bp in length
* gap of unknown length
* 150499 160979: contig of 10481 bp in length
* gap of unknown length
* 160980 163035: contig of 2056 bp in length.
FEATURES
Location/Qualifiers
1..163035
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="hRPK.107 N.19"
/clone_lib="RPC1-11 human BAC library"

ORIGIN
Alignment Scores:
Pred. NO.: 5.39e+03 Length: 163035
Score: 103.00 Matches: 38
Percent Similarity: 37.74% Conservative: 2
Best Local Similarity: 35.85% Mismatches: 33
Query Match: 22.59% Indels: 34
DB: 2 Gaps: 5

US-09-972-032-2 (1-79) x AC006405 (1-163035)
Qy 3 GlyArgProArgValSerAlaGlyCysGlyPhe-----AlaAspAlaHisTrr 19
Db 145698 GTTCGGCCGGGGGTGTAGGGGAGCGGGGGCGTTTCGCATCTGACCGGGGCCAC---- 145642
Qy 20 ThrGlyLeuTrrThrGlyLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 39
Db 145641 ---GGGGCATGTGACGGGC---GGCCGGGGCATGTGAGGGGGCCACGGGAGGACAA 145588
Qy 40 AlaSerProThrProaspCysAlaSerArgTrrProArgSerAlaSerArgTrrProTrr 59
Db 145587 CCGGGCGGACGGCTCTCTAGGGCGCCCGGGGGCCCTCGG---AGGCCAAGTGGACCGG- 145532
Qy 60 SerAlaGlyLeuThrValArgAspArgProGln----- 70

```

Db 145531 AGGAGGCGGGGACAGCGAGGCGGAGCCAGCGGCTCAGGCGAGGAGA 145472
Qy 71 -----LeuGlyGlu 73
Db 145471 GCTGCGCGGCGACTGGGCGAGCGGAGCGGCGAGCGCGCTGGAGTGTGTGGGTCN 145412
Qy 74 LeuCyseMetGlyArgGly 79
Db 145411 CTCTGCGCGGCGCGAGGG 145394
RESULT 5
S53698
LOCUS
DEFINITION
S53698
VERSION
S53698.1 GI:234760
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
Bernard, O., Lecointe, N., Jonveaux, P., Souyri, M., Mauchauffe, M.,
Berger, R., Larsen, C.J. and Mathieu-Mahul, D.
TITLE
Two site-specific deletions and t(1;14) translocation restricted to
human T-cell acute leukemias disrupt the 5' part of the tal-1 gene
JOURNAL
Oncogene 6 (8), 1477-1488 (1991)
MEDLINE
91360285
PUBMED
1886719
REMARK
GenBank staff at the National Library of Medicine created this
entry [NCBI gibbsq 53698] from the original journal article.
COMMENT
Exon 1a (nt 269-379) and exon 1b (nt 657-890) are involved in
alternative splicing of tal-1 gene.
FEATURES
source
1..900
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
gene
1..900
/gene="tal-1"
/note="exons 1a, 1b and flanking DNA"
ORIGIN
Alignment Scores:
Pred No.: 118 Length: 900
Score: 100.00 Matches: 28
Percent Similarity: 41.25% Conservative: 5
Best Local Similarity: 35.00% Mismatches: 31
Query Match: 21.93% Indels: 16
DB: 9 Gaps: 4
US-09-972-032-2 (1-79) x S53698 (1-900)
Qy 4 ArgProArgArgValSerAlaGlyCys-----GlyPheAlaAsp 16
Db 195 CGGCGCGCGCGCGCTCGCGCCCATGCTAACGAGCGCTGGAACGCTGAGTGGGATTACAGCG 254
Qy 17 AlaHisTrpThrGlyLeuTrpThrGlyLeuGlyGlnGlyGlyGlyPro 36
Db 255 CGTGGTGGAGGCGCGG---CGCGGCGCGGCGCGCGGCGCGGCGCGCGG--- 308
Qy 37 GluGlyGlnAlaSerProThrProAspCysAlaSerArgTrpProArgSerAlaSerArg 56
Db 309 AGGGGCGGCGCGCGCGCGCTCAGACCGGGCGCTCAAAATGGCCACACGCGGTACCCCGCT 368
Qy 57 -----TTPProTrpSerIaGlyLeuThrValArgAspArgPro 69
Db 369 AGCGGAAAAACGTGAGCATTTCTGCGCTTTTCTAGGGGAAAAGCAACCGCGCGCTCCCA 428
RESULT 6

AC006970
LOCUS
DEFINITION
AC006970
VERSION
AC006970.6 GI:10312289
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS
Sulston, J.E. and Wilson, R.
TITLE
Toward a complete human genome sequence
JOURNAL
Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE
99063792
PUBMED
9847074
REFERENCE
AUTHORS
Du, H., Haakenson, B. and Stoneking, T.
TITLE
The sequence of Homo sapiens PAC clone RP4-725G10
JOURNAL
Unpublished (2001)
REFERENCE
AUTHORS
Waterston, R.H.
TITLE
Direct Submission
JOURNAL
Submitted (05-MAR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS
Waterston, R.H.
TITLE
Direct Submission
JOURNAL
Submitted (27-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS
Wilson, R.
TITLE
Direct Submission
JOURNAL
Submitted (15-OCT-2003) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
On Sep 27, 2000 this sequence version replaced gi:9838025.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu>
Contact: sapiens@watson.wustl.edu
----- Summary Statistics

Center project name: H_DJ0725G10

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu>

SOURCE INFORMATION:
This clone was derived from human PAC library RPI-4, prepared by Pieter de Jong and coworkers at <http://www.chori.org> using the


```

QY 36 -----ProGluGlyGlnAlaSerProThrPr 44
Db 80476 TTGCTGGCGGGGGTGGCCACACATAGCCACCCGAGGGT-----GGTCCCCCGA 80529
QY 44 cAspCysAlaSerArgTrrProArgSerAlaSerArgTrrProTrrSerAlaGlyLeuTh 64
Db 80530 GGGCGCTCTCCAGGGTGGTCCAGGCTAGCTCGGACCTGGCAGTGGGGAGTGGCGAGA 80589
QY 64 rValArg-----AspArgPr 69
Db 80590 GCTGGCTGGAAAGCGGGGGCGCTGGCTGGCGAAGAACTAGCTGCTGACACGCC 80649
QY 69 oGlnLeuGlyGluLeuCysMetGlyArggly 79
Db 80650 GCAAGTACCACGACCTGCTCGCGCTCGGC 80680

RESULT 7
AF164165/c 2163 bp DNA linear VRT 13-MAR-2000
LOCUS Certhidea olivacea MHC class IIB antigen gene, partial cds.
DEFINITION
ACCESSION AF164165
VERSION AF164165.1 GI:6941861
KEYWORDS
SOURCE Certhidea olivacea
ORGANISM Certhidea olivacea
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Passeriformes; Passeroidea;
Fringillidae; Emberizinae; Certhidea.
REFERENCE
Sato,A., Figueroa,F., Mayer,W.E., Grant,P.R., Grant,R. and Klein,J.
MHC class II genes of Darwin's Finches: Divergence by point
mutations and reciprocal recombination
(in) Kasahara,M. (Ed.);
MAJOR HISTOCOMPATIBILITY COMPLEX. EVOLUTION, STRUCTURE, AND
FUNCTION: 518-541;
Springer-Verlag, Tokyo, Japan (2000)
REFERENCE
Sato,A., Figueroa,F., Mayer,W.E., Grant,P.R., Grant,R. and Klein,J.
Direct Submission
TITLE Submitted (30-JUN-1999); Immunogenetics, Max-Planck-Institute for
Biology, Corrensstr. 42, Tuebingen D-72076, Germany
JOURNAL
FEATURES
source
1..2163
/organism="Certhidea olivacea"
/mol_type="genomic DNA"
/isolate="D1103-I2"
/db_xref="taxon:48880"
join(<1..223,2052..>2163)
/product="MHC class IIB antigen"
join(<1..223,2052..>2163)
/notes="group 3-2"
/codon_start=1
/product="MHC class IIB antigen"
/protein_id="AAF32250.1"
/db_xref="GI:6941866"
/taxname="TEKRVYVREIYNRLMVFDSDVGHVGTPTPGERVAKHNSD
PAILEDRAQVDICRHYVSVPFITERVPSVISLVPSSSQGPFPALLRDG
FLPCDPGEV"
<1..223
/number=2
2052..>2163
/number=3

```

```

ORIGIN
Alignment Scores:
Pred. No.: 307 Length: 2163
Score: 98.50 Matches: 34
Percent Similarity: 32.23% Conservative: 5
Best Local Similarity: 28.10% Mismatches: 27
Query Match: 21.60% Indels: 55
DB: 5 Gaps: 6

```

```

US-09-972-032-2 (1-79) x AF164165 (1-2163)
QY 5 ProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrrThrGlyLeu----- 22
Db 430 CTAGGCGATGACTGGAGGGTTTGGATGACTGAGAGGCACCTGGAAGGGAGTGGAGAG 371
QY 23 -----TrrThrGlyLeuGlyGluGlyGlnGlu 31
Db 370 ACTGGGATGATTGGATGAGTGGAGTGGAGTGGAGGAGGAGTGGAGTGGAG 311
QY 32 Gly-----GlyIleGly-----Pro 36
Db 310 GCATGRRAGGVGNRRTTMGACTGGGATTCGTGGGATGGTCACTGCCAGGGCAGGCC 251
QY 37 GluGlyGlnAlaSer----- 41
Db 250 GAGGGCACACGCTCTGCCCGCGCTCACCTCGCGCGCTCGATGATGAACGGGGGACACCT 191
QY 42 -----ProThrProAspCysAlaSerArgTrrPro----- 51
Db 190 CGTAGTTCGCGCGAGAGCGTGCACCTGAGCGCGTCTGCTCCAGTATGCCCGGT 131
QY 52 ArgSerAlaSerArgTrrProTrrSer-----AlaGlyLeuThrValArgAspArgPro 69
Db 130 CGCTGTTCCAGTCTTGGCCACCTCTCCCATAGGGGGTGAACCCACGTAAGTGCCTCA 71
QY 70 Gln 70
Db 70 COT 68

RESULT 8
AC092012 153675 bp DNA linear HTG 13-JUN-2001
LOCUS Felis catus clone RP86-89F20, WORKING DRAFT SEQUENCE, 5 unordered
DEFINITION
ACCESSION AC092012
VERSION AC092012.1 GI:14389334
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Felis catus (cat)
ORGANISM Felis catus
REFERENCE
1 (bases 1 to 153675)
AUTHORS
Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L.,
Granite,S., Guan,X., Gupta,J., Ho,S.L., Idol,J.R., Karlins,E.,
Lee-Lin,S.-Q., Legaspi,R., Lim,M., Maduro,Q.L., Maduro,V.B.,
Masello,C., Mastrian,S.D., McCloskey,J.C., McDowell,J.,
Pearson,R., Prasad,A., Shevchenko,Y., Snyder,B., Stantropop,S.,
Thomas,J.W., Thomas,P.J., Touchman,J.W., Tsurgon,C., Vogt,J.L.,
Walker,M.A., Wetherby,K.D., Zhang,L.-H. and Green,E.D.
NISC Comparative Sequencing Initiative
Unpublished
TITLE
JOURNAL
REFERENCE 2 (bases 1 to 153675)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (13-JUN-2001) NIH Intramural Sequencing Center, 8717
Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT
-----Genome Center Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@nih.gov
-----Project Information
Center project name: cfb
Center clone name: 089F20
-----Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 150701 bases at least Q40
Consensus quality: 151335 bases at least Q30
Consensus quality: 151627 bases at least Q20

```

```

Insert size: 144000; agarose-fp
Insert size: 153275; sum-of-contigs
Quality coverage: 13.08x in Q20 bases; agarose-fp
Quality coverage: 12.28x in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 12153: contig of 12153 bp in length
* 12154 12253: gap of unknown length
* 12254 12254: contig of 10195 bp in length
* 22449 22448: gap of unknown length
* 22549 46085: contig of 23537 bp in length
* 46086 46185: gap of unknown length
* 46186 72533: contig of 26348 bp in length
* 72534 72534: gap of unknown length
* 72634 153675: contig of 81042 bp in length.
*
* Location/Qualifiers
*   1..153675
*     /organism="Felis catus"
*     /mol_type="genomic DNA"
*     /db_xref="taxon:9685"
*     /clone="RP86-89P20"
*     /clone_lib="RP86"
*   1..12153
*     /notes="assembly_fragment"
*   12254..22448
*     /notes="assembly_fragment"
*     /clone_end:SP6
*     vector_side:right
*   22549..46085
*     /notes="assembly_fragment"
*     clone_end:T7
*     vector_side:right
*   46186..72533
*     /notes="assembly_fragment"
*   72634..153675
*     /notes="assembly_fragment"
*
* ORIGIN
*
Alignment Scores:
Pred. No.: 1.04e+04 Length: 153675
Score: 98.50 Matches: 30
Percent Similarity: 31.48% Conservative: 4
Best Local Similarity: 27.78% Mismatches: 27
Query Match: 21.60% Indels: 47
DB: 2 Gaps: 5

US-09-972-032-2 (1-79) x AC092012 (1-153675)
QY 19 TrpThrGlyLeuTrpThrGlyLeuGlyGlu-----GlyGlnGluGlyGly 33
Db 153038 TGGTCCTCCCTCTGGCAGCGGGAGGAGGTCGCCCTCGGGGATCG 153097
QY 34 IleGlyPro-----GluGlyGlnAlaSerProThrProAsp 45
Db 153098 CTGCCCCGAGTACTCCAGATATTTGGGGAGGGAGTGGGAGAGGCTGCAGT 153157
QY 46 Cys-----AlaSerArgTrpProArgSerAlaSer 55
Db 153158 TGTCTGGCAGCAAGAGGAGCAGCAGTAGAGGGGCTTCGTTTGGAAATCGGAGGCTGCG 153217
QY 56 ---ArgTrpProTrpSerAla----- 61
Db 153218 TGCGGTGGCCCTGGCCGCTCTCCGTCGTCGAGACCGGCTTGGGGGGGAATCAG 153277
QY 62 -----GlyLeuThrValArgAspArgProGlnLeu 71
Db -----

```

```

Db 153278 GCCCTGGAAGAGGAGTCCGCGGTCCGCGGGGCAAGGTTCCGATCGGGGAGGCG 153337
QY 72 GlyGluLeuCysMetGlyArgGly 79
Db 153338 GCGCGCTGTGGACGGGCGAGGGC 153361

RESULT 9
AC134385 185997 bp DNA linear HTG 27-NOV-2003
LOCUS Papio anubis clone rp41-10512, WORKING DRAFT SEQUENCE, 2 ordered
DEFINITION pieces.
ACCESSION AC134385
VERSION AC134385.10 GI:38564357
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT.
SOURCE Papio anubis (olive baboon)
ORGANISM Papio anubis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
Cercopithecinae; Papio.
REFERENCE 1 (bases 1 to 185997)
AUTHORS Commen,S., Shaikh,T. and Roe,B.A.
TITLE Papio anubis BAC Clone rp41-10512
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 185997)
AUTHORS Commen,S. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (26-SEP-2002) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 185997)
AUTHORS Commen,S., Shaikh,T. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-2003) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 118468: contig of 118468 bp in length
* 118469 118568: gap of unknown length
* 118569 185997: contig of 67429 bp in length.
*
* Location/Qualifiers
*   1..185997
*     /organism="Papio anubis"
*     /mol_type="genomic DNA"
*     /db_xref="taxon:9555"
*     /clone="rp41-10512"
*     /clone_lib="RPCI - 41 Male (Olive) Baboon BAC Library"
*
ORIGIN
Alignment Scores:
Pred. No.: 1.32e+04 Length: 185997
Score: 98.00 Matches: 25
Percent Similarity: 50.00% Conservative: 7
Best Local Similarity: 39.06% Mismatches: 20
Query Match: 21.49% Indels: 12
DB: 2 Gaps: 4

US-09-972-032-2 (1-79) x AC134385 (1-185997)
QY 14 PheAlaAspAlaHisTrpThrGlyLeuTrpThrGlyLeuGlyGlnGlyGly 33

```

```

Db 165264 TTCTCAGATTTCCTGCGCTGGATGGTGGGAGGAGGAGGTGGTGGTGCAGCTGCCTGTC 165205
QY 34 lileglyProGluGlyGlnAlaSerProThrProAspCysAlaSerArgTrpProArgSer 53
Db 165204 CTAGGCGCTG-----AGGCGCTTCACCACTCCCTCCCTGTCG-----CCAAGGAAG 165160
QY 54 AlaSerArg-----TTPProTrpSerAlaGlyLeuThrValArgAspArgProGln 70
Db 165159 GCAGTCAGACCGCGGCGCTGCTGCTGTCATCAGGATG-----GGTCCCTCT 165112
QY 71 LeuGlyGluLeu 74
Db 165111 GCTGGGAAGATG 165100

RESULT 10
AC138018 191292 bp DNA linear HTG 11-SEP-2003
LOCUS Papio anubis clone rp41-11lp21, WORKING DRAFT SEQUENCE, 7 ordered
DEFINITION pieces.
AC138018
AC138018.6 GI:34576501
VERSION HTG: HTGS PHASE2; HTGS_DRAFT.
KEYWORDS Papio anubis (Olive baboon)
SOURCE Papio anubis
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Cercopitheciinae; Papio.
1 (bases 1 to 191292)
Li,J., Yang,L., Lau,C.C.Y., Shaikh,T. and Roe,B.A.
Papio anubis BAC Clone rp41-11lp21
Unpublished
2 (bases 1 to 191292)
Li,J., Yang,L., Lau,C.C.Y., Shaikh,T. and Roe,B.A.
Direct Submission
Submitted (10-DEC-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 520 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 191292)
Li,J., Yang,L., Lau,C.C.Y., Shaikh,T. and Roe,B.A.
Direct Submission
Submitted (11-SEP-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 520 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Sep 11, 2003 this sequence version replaced gi:30270648.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 8554: contig of 8554 bp in length
8555 8654: gap of unknown length
8656 64329: contig of 55675 bp in length
64330 64429: gap of unknown length
64430 67419: contig of 2990 bp in length
67420 67519: gap of unknown length
67520 78467: contig of 10948 bp in length
78468 78567: gap of unknown length
78568 90992: contig of 12425 bp in length
90993 91092: gap of unknown length
91093 186191: contig of 95099 bp in length
186192 186291: gap of unknown length
186292 191292: contig of 5001 bp in length.
Location/Qualifiers

```

```

source
1..191292
/organism="Papio anubis"
/mol_type="genomic DNA"
/db_xref="taxon:9555"
/clone="rp41-11lp21"
/clone_lib="RPC1 - 41 Male (Olive) Baboon BAC Library"

ORIGIN
Alignment Scores:
Pred. No.: 1.35e+04 Length: 191292
Score: 98.00 Matches: 25
Percent Similarity: 50.00% Conservative: 7
Best Local Similarity: 39.06% Mismatches: 20
Query Match: 21.49% Indels: 12
DB: 2 Gaps: 4

US-09-972-032-2 (1-79) x AC138018 (1-191292)
QY 14 PheAlaAspAlaHisTrpThrGlyLeuThrTrpGlyLeuGlyGlnGluGlyGly 33
Db 46469 TTCTCAGATTTCCTGCGCTGGATGGTGGGAGGAGGAGGTGGTGGTGCAGCTGCCTGTC 46528
QY 34 lileglyProGluGlyGlnAlaSerProThrProAspCysAlaSerArgTrpProArgSer 53
Db 46529 CTAGGCGCTG-----AGGCGCTTCACCACTCCCTCCCTGTCG-----CCAAGGAAG 46573
QY 54 AlaSerArg-----TTPProTrpSerAlaGlyLeuThrValArgAspArgProGln 70
Db 46574 GCAGTCAGACCGCGGCGCTGCTGCTGTCATCAGGATG-----GGTCCCTCT 46621
QY 71 LeuGlyGluLeu 74
Db 46622 GCTGGGAAGATG 46633

RESULT 11
SC0939131 303550 bp DNA linear BCT 11-FEB-2003
LOCUS Streptomyces coelicolor A3(2) complete genome; segment 28/29.
DEFINITION Streptomyces coelicolor A3(2) complete genome; segment 28/29.
ACCESSION AL939131 AL049863 AL158057 AL158060 AL158061 AL355753 AL356592
AL356932 AL357613 AL590435 AL592126 AL596248 AL645882
AL939131.1 GI:24418961
VERSION Streptomyces coelicolor A3(2)
KEYWORDS Streptomyces coelicolor A3(2)
SOURCE Streptomyces coelicolor A3(2)
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
AUTHORS Bentley, S.D., Chater, K.F., Cerdeno-Tarraga, A.M., Challis, G.L.,
Thomson, N.R., James, K.D., Harris, D.E., Quail, M.A., Kieser, H.,
Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M.,
Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S.,
Huang, C.H., Kieser, T., Larke, L., Larke, L., Murphy, L., Oliver, K., O'Neill, S.,
Rabinowitz, E., Rajandream, M.A., Rutherford, K., Rutter, S.,
Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S.,
Taylor, K., Warren, T., Wietzorrek, A., Woodward, J., Barrett, B.G.,
Parkhill, J. and Hopwood, D.A.
Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)
JOURNAL Nature 417 (6885), 141-147 (2002)
MEDLINE 21996410
PUBMED 12000953
REFERENCE 2 (bases 1 to 303550)
Bentley, S.D.
Direct Submission
Submitted (09-MAY-2002) Submitted on behalf of the Streptomyces
sequencing team, Sanger Institute, Wellcome Trust Genome Campus,
Hinxton, Cambridge CB10 1SA E-mail: sdb@sanger.ac.uk
On or before Oct 29, 2002 this sequence version replaced
gi:20520692, gi:20520799, gi:20520800, gi:20520856, gi:20520802,
gi:20520783, gi:20520865, gi:20520870, gi:20520878.
Location/Qualifiers
1..303550
/organism="Streptomyces coelicolor A3(2)"
source

```

```
/mol_type="genomic DNA"
/strain="A3(2)"
/db_xref="taxon:100226"
135..1643
/gene="SC07252"
/notes="synonym: SC5H1.40c"
CDS
135..1643
/gene="SC07252"
/notes="SC5H1.40c, possible regulatory protein, len: 502
aa; similar to e.g. SW:SP15_STRGR (EMBL:M32687)
Streptomyces griseus sporulation protein (529 aa), fasta
scores; opt: 394 z-score: 452.2 E(): 7.2e-18, 27.9%
identity in 530 aa. Similar to SC7A1.26 (EMBL:AL034447)
S.coelicolor possible transcriptional regulator (500 aa)
(43.0% identity in 491 aa overlap)"
/codon_start=1
/transl_table=11
/product="putative regulatory protein"
/protein_id="CAB42965.1"
/db_xref="GI:4835343"
/db_xref="SPTREMBL:Q9X7U7"
/translation="MSKPNTRLSDLFGLAGWSKGLARLVNQAAMGHPOLATDTS
RVRWIDMGIPIPDVPRVLAALFTELRGEVVTIEDLGLVHGRAGKRPDSESEHPD
GVWPAPRTAAVLTEFTGMDLMLNRLGLVGAGALAAGSALSAMDMWLTDPALAAD
APDHLQHPADGPDYEAAPIGSQSEVESLEKSVFRAWDAGGGQKRAVVGQL
NEVGMLAYHPHPLQRLWGAANAVLAGWMSHDVLEPTAQKVFVIAAAREGG
DRPAGALSRAARQVHLGKLPDEALDLMLKQSGSGEQLPRTKMLTYIEAWAQS
MGKQAMRTLGRAEDLVSKADVPDPMMQOTFKSEDLYGQALAYRTLAFEPGAA
AAACHYADKALARI DGRORSKIFDYLMSACFIADDPQADRYARLARLVMSGNS
QRTWDLRQMYRLTAEGYKPKIQELREELKALPKGSRPGKPGGTAPA"
/complement(1663..2979)
/gene="SC07253"
/notes="synonym: SC5H1.39"
complement(1663..2979)
/gene="SC07253"
/notes="SC5H1.39, hypothetical protein, len: 438 aa;
unknown function, probable CDS suggested by positional
base preference, GC frame analysis and amino acid
composition"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB42964.1"
/db_xref="GI:4835342"
/db_xref="GOA:Q9X7U6"
/db_xref="SPTREMBL:Q9X7U6"
/translation="MSSHLSADHPAAQSQPGQSVREALISQTRLRGDDVAVRRSDQD
TTDPRGWQALYDLALHQLTDLDAHLAQLRDGPAPAPRAGSLLSRVGAENWLLTDE
AENGAELPHILGRDPASVPLSLDELPSVLTEDRPKLTAMVTCLVDGRPIDGEFRV
RPGGSVTVMMGCEPVLDAQSGTASMAWLVLDGRRSORAVSETRDSLRRPASPM
HGLLRFPDEGRPDLGALHLPSTATGRSREGWYDACPGLGAGETLITVGDGFLP
EADYAGPVLIGALRGALAGTAPRGLRSLAEELLRATALAPVDGAVYGYRPDHTL
TWAHSGAPLPIFGGTGHALSGPDGAGQDGPQSGAEATLEAGDLLLLHTGALEPEA
VEHLITLAPRITADSGDGVRLVITRELDGPPGRQACVLIARVTP"
/complement(3252..4280)
/gene="SC07254"
/notes="synonym: SC5H1.38"
complement(3252..4280)
/gene="SC07254"
/notes="SC5H1.38, probable myo-inositol dehydrogenase, len:
342 aa; similar to many e.g. SW:M12D_BACSU (EMBL:M76431),
idh, Bacillus subtilis myo-inositol 2-dehydrogenase (344
aa), fasta scores; opt: 864 z-score: 958.5 E(): 0, 38.1%
identity in 336 aa overlap. Also similar to SW:STRI_STRGR
(EMBL:Y00459), stri, Streptomyces griseus streptomycin
biosynthesis myo-inositol oxidoreductase (348 aa) (32.7%
identity in 343 aa overlap). Weakly similar to TR:O69945
(EMBL:AL023862) S.coelicolor possible oxidoreductase (430
aa) (29.8% identity in 228 aa overlap). Contains Pfam
match to entry PF01408 GFO IDH Moca, Oxidoreductase
family, score 153.50, E-value 3.6e-42"
/codon_start=1
/transl_table=11
/product="putative myo-inositol dehydrogenase"
/protein_id="CAB42963.1"
/db_xref="GI:4835341"
/db_xref="GOA:Q9X7U5"
/db_xref="SPTREMBL:Q9X7U5"
/translation="MSELLGVAVLGAHGMDHRRVQVVVGARVAADPDABRAK
EAVGIGGTGRITVHTDVEALDAPGVEAVLIASPGEAHEEALAAFGALPVLCERP
MAPNSAGALRVVEAEARLGRRLAQIGFMRRYDAEYRQLKSLDGLGRPLMLRCHVR
NVSPPHFTSAMLINSVSHIEDAARLLGQELSAVTVLRPRPSAGAPEGLLDQLVL
PETGGAVDVEVFNCGFGYVRCVCAVCRAGSARIGAAHTMTVTAAGAAEEVPDY
LVRDADYDREVQSWDATRGLVTGPGTWGYYAAAAGVAGVRLDTGVTGRTVDMAP
RPSLHURA"
/complement(3549..4268)
/gene="SC07254"
/notes="Pfam match to entry PF01408 GFO IDH Moca,
Oxidoreductase family, score 153.50, E-value 3.6e-42"
complement(4291..4295)
/notes="possible RBS"
complement(4404..5498)
/gene="SC07255"
/notes="synonym: SC5H1.37"
complement(4404..5498)
/gene="SC07255"
/notes="SC5H1.37, hypothetical protein, len: 364 aa;
unknown function, similar to TR:P73120 (EMBL:D90903)
Synchocystis sp. hypothetical protein (358 aa), fasta
scores; opt: 492 z-score: 559.6 E(): 7.6e-24, 31.6%
identity in 364 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAB42962.1"
/db_xref="GI:4835340"
/db_xref="SPTREMBL:Q9X7U4"
/translation="MEHDLSPATLGLRRPRPYPAVSVLTPTTHREPDNAQRVLRN
AVAAKQLEEDPAVDRRAEVSLELDRAELDLTVAEDGLAIFAPGHEHVQWTLS
RVSPPRVLSDTFLTRNLVAAQAERPFWLSVAADVKLMSGGAGVVEHLGGFPL
DRPAEDFPERQERIGDAPSTFRDEGRRLRDTDAMGRILRAHPRLVYUTGPEPL
SLFEAGALASDAVTPHGLAGHTAEAWQVVRPVLKREARGSVAEVAELDARGR
KDFAGVDLWESAVTGRVLLAEENFKVTMRDDGHELLFADDDGLDAREDIV
EQCLTGDVRFVDPDGLGDVDDGIAGVLYR"
/complement(5513..5516)
/notes="possible RBS"
complement(5580..7211)
/gene="SC07256"
/notes="synonym: SC5H1.36"
complement(5580..7211)
/gene="SC07256"
/notes="SC5H1.36, probable protease (putative secreted
protein), len: 543 aa; similar to many e.g. TR:Q54398
(EMBL:U42758), SlpD, Streptomyces lividans protease, (539
aa), fasta scores; opt: 697 z-score: 769.7 E(): 0.35.8%
identity in 545 aa overlap. Similar to other S.coelicolor
proteases e.g. SlpD, SCH5.02c (EMBL:AL03563) proteinase
(539 aa) (35.8% identity in 545 aa overlap). Contains
possible N-terminal signal sequence. Contains Pfam match
to entry PF00561 abhydrolase, alpha/beta hydrolase fold,
score 55.00, E-value 1.6e-12"
/codon_start=1
/transl_table=11
/product="putative protease (putative secreted protein)"
/protein_id="CAB42961.1"
/db_xref="GI:4835339"
/db_xref="GOA:Q9X7U3"
/db_xref="SPTREMBL:Q9X7U3"
/translation="MQAKLETRPSVRCNVTGALGLALIGALPAASGATTAATRTDT
SGSLRVFDQKRVETACGETEMPKQLQCKGVTFUDTSRPRGGTLDVALARYATGDS
RGSVLNPPGGSGINELALGKGFHMLTNGYVDVTFDPGRGDSFVSCGPAILKI
MEATDYGCTMDTPPEGLERLDAAECAKSGSPVLPHGTVAADARDVMRRALGDDR
LNYLFGSYGTRLGAVYARFPDKVGRVLDGVDLTLEPLAQGLAGAGQOATLENFL
DWCVEDVACPFGQDARDQVERLVASLSDPVPFSAFPPFTGQDMVGATGQALYSR
ELMPSLERALAQLEDGDTGRLGEGFSSGGVTFPVRAAVNAPVRGDTAGLTDEBDVPM
DNLPAALMAINCADDPRPTAAQVTASISDLRLARYEDVSPVFGVRLTQVLMCYGRPR
```


GTDIRDDVKDLTARVLLVCTRGDPATPYRTWTETADRLGPSAVLNDNRGEGHTGYA
SSKCVHRKVDVFLYGLSPDGGSSCGFESTGDDGG"
complement(5910. .6809)
/gene="SC07256"
/note="Pfam match to entry PF00561 abhydrolase, alpha/beta
hydrolase fold, score 55.00, E-value 1.6e-12"
complement(7217. .7220)
/note="possible RBS"
7437. .7440
/note="possible RBS"
7445. .8158
/gene="SC07257"
/note="synonym: SC5H1.35c"
7445. .8158
/gene="SC07257"
/note="SC5H1.35c, possible secreted protein, len: 237 aa;
unknown function, probable CDS suggested by positional

Alignment Scores:
Pred. No.: 1.98e+04 Length: 303550
Score: 98.00 Matches: 26
Percent Similarity: 40.79% Conservative: 5
Best Local Similarity: 34.21% Mismatches: 35
Query Match: 21.49% Indels: 10
DB: 1 Gaps: 3

US-09-972-032-2 (1-79) x SC0939131 (1-303550)

QY 4 ArgProArgValSerAlaGly-----CysGlyPheAlaAspAlaHisTrpThr 20
||||| :|||
Db 80456 CGGCGGAGCGGTACACCGGAGGTGCGGACATGCTCGCACCGGACGGTCTCTGGAGC 80397
||||| :|||

QY 21 GlyLeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGly 40
||||| :|||
Db 80396 ACGGATGGTCTACTTCGACGTCGCGTCTCCAGCGGTATCCGACGGTTCGAGTTCGGG 80337
||||| :|||

QY 41 SerPro-----ThrProAspCysAlaSerArgTrpProArgSerAlaSerArg 56
||||| :|||
Db 80336 TCGCGGAGCTGCTGCTGCGGCTCCACCGCGTCTGCTGCGCGCTCGCGCGCGC 80277
||||| :|||

QY 57 TrpProTrp-----SerAlaGlyLeuThrValArgAspArgPro 69
||||| :|||
Db 80276 TGGTCGACGCGCGCGGAGTGGCGCGCGCGGCGGAGAACGCGCG 80229
||||| :|||

RESULT 12
BD179739 1476 bp DNA linear PAT 15-MAY-2003
LOCUS
DEFINITION Highly thermophilic bacterium-derived protein and gene encoding it.
ACCESSION BD179739
VERSION BD179739.1 GI:30790657
KEYWORDS JP 2002325574-A/230.
SOURCE Thermus thermophilus
ORGANISM Thermus thermophilus
Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
Thermus.
REFERENCE 1 (bases 1 to 1476)
AUTHORS Kuramitsu,N. and Yokoyama,S.
TITLE Highly thermophilic bacterium-derived protein and gene encoding it
JOURNAL Patent: JP 2002325574-A 230 12-NOV-2002;
THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH
COMMENT OS Thermus thermophilus
PN JP 2002325574-A/230
PD 12-NOV-2002
PF 23-FEB-2001 JP 2001116171
PI NARUKI KURAMITSU,SHIGEYUKI YOKOYAMA
PC C12N15/09,C12N15/09,C07K14/195,C12N1/15,C12N1/19,C12N1/21, PC
C12N5/10,
PC C12N9/88,C12P21/02/(C12N9/88,C12R1/01),(C12N15/09,C12R1/01),
PC (C12P21/02,C12R1/01),C12N15/00,C12N15/00,C12N5/00,
CC C12R1/01
CC Highly thermophilic bacterium-derived protein and gene CC
encoding it
FH Key Location/Qualifiers

misc_feature
complement(5910. .6809)
/gene="SC07256"
/note="Pfam match to entry PF00561 abhydrolase, alpha/beta
hydrolase fold, score 55.00, E-value 1.6e-12"
complement(7217. .7220)
/note="possible RBS"
7437. .7440
/note="possible RBS"
7445. .8158
/gene="SC07257"
/note="synonym: SC5H1.35c"
7445. .8158
/gene="SC07257"
/note="SC5H1.35c, possible secreted protein, len: 237 aa;
unknown function, probable CDS suggested by positional

Alignment Scores:
Pred. No.: 1.98e+04 Length: 303550
Score: 98.00 Matches: 26
Percent Similarity: 40.79% Conservative: 5
Best Local Similarity: 34.21% Mismatches: 35
Query Match: 21.49% Indels: 10
DB: 1 Gaps: 3

US-09-972-032-2 (1-79) x SC0939131 (1-303550)

QY 4 ArgProArgValSerAlaGly-----CysGlyPheAlaAspAlaHisTrpThr 20
||||| :|||
Db 80456 CGGCGGAGCGGTACACCGGAGGTGCGGACATGCTCGCACCGGACGGTCTCTGGAGC 80397
||||| :|||

QY 21 GlyLeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGly 40
||||| :|||
Db 80396 ACGGATGGTCTACTTCGACGTCGCGTCTCCAGCGGTATCCGACGGTTCGAGTTCGGG 80337
||||| :|||

QY 41 SerPro-----ThrProAspCysAlaSerArgTrpProArgSerAlaSerArg 56
||||| :|||
Db 80336 TCGCGGAGCTGCTGCTGCGGCTCCACCGCGTCTGCTGCGCGCTCGCGCGCGC 80277
||||| :|||

QY 57 TrpProTrp-----SerAlaGlyLeuThrValArgAspArgPro 69
||||| :|||
Db 80276 TGGTCGACGCGCGCGGAGTGGCGCGCGCGGCGGAGAACGCGCG 80229
||||| :|||

RESULT 12
BD179739 1476 bp DNA linear PAT 15-MAY-2003
LOCUS
DEFINITION Highly thermophilic bacterium-derived protein and gene encoding it.
ACCESSION BD179739
VERSION BD179739.1 GI:30790657
KEYWORDS JP 2002325574-A/230.
SOURCE Thermus thermophilus
ORGANISM Thermus thermophilus
Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
Thermus.
REFERENCE 1 (bases 1 to 1476)
AUTHORS Kuramitsu,N. and Yokoyama,S.
TITLE Highly thermophilic bacterium-derived protein and gene encoding it
JOURNAL Patent: JP 2002325574-A 230 12-NOV-2002;
THE INSTITUTE OF PHYSICAL AND CHEMICAL RESEARCH
COMMENT OS Thermus thermophilus
PN JP 2002325574-A/230
PD 12-NOV-2002
PF 23-FEB-2001 JP 2001116171
PI NARUKI KURAMITSU,SHIGEYUKI YOKOYAMA
PC C12N15/09,C12N15/09,C07K14/195,C12N1/15,C12N1/19,C12N1/21, PC
C12N5/10,
PC C12N9/88,C12P21/02/(C12N9/88,C12R1/01),(C12N15/09,C12R1/01),
PC (C12P21/02,C12R1/01),C12N15/00,C12N15/00,C12N5/00,
CC C12R1/01
CC Highly thermophilic bacterium-derived protein and gene CC
encoding it
FH Key Location/Qualifiers

FT CDS Location/Qualifiers
1..1476
/organism="Thermus thermophilus"
/mol_type="genomic DNA"
/db_xref="taxon:274"

ORIGIN
Alignment Scores:
Pred. No.: 262 Length: 1476
Score: 97.50 Matches: 22
Percent Similarity: 42.31% Conservative: 0
Best Local Similarity: 42.31% Mismatches: 23
Query Match: 21.38% Indels: 7
DB: 6 Gaps: 2

US-09-972-032-2 (1-79) x BD179739 (1-1476)

QY 19 TrpThrGlyLeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGly 38
||||| :|||
Db 518 TGGACCGGCTTCTGGAGGCGCTTTGGACCGGATCGGCGCTGGCGGGCGGAGGGG 577
||||| :|||

QY 39 GlnAlaSerProThrProAspCysAlaSerArgTrpProArgSerAlaSerArgTrpPro 58
||||| :|||
Db 578 GTGGGGTCC-----TCCTTTTGGACGAGGAGGCGCTTCTCTCTGGCCC 619
||||| :|||

QY 59 ---TrpSerAlaGlyLeuThrValArgAspArgPro 69
||||| :|||
Db 620 GGGTGTCCAGGCGCTTGCCTTCCTCCACCCCT 655
||||| :|||

RESULT 13
AC006123 45858 bp DNA linear PRI 17-DEC-1998
LOCUS
DEFINITION Homo sapiens chromosome 19, cosmid R28177, complete sequence.
ACCESSION AC006123
VERSION AC006123.1 GI:3970939
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 45858)
AUTHORS Lamerdin,J.E., McCready,P.M., Skowronski,E., Viswanathan,V.,
Burkhardt-Schultz,K., Gordon,L., Dias,J., Kyle,A., Brower,A.,
Stilwagen,S., Phan,H., Velasco,N., Do,L., Regala,W., Terry,A.,
Garnes,J., Danganan,L., Erler,A., Christensen,M., Georgescu,A.,
Avila,J., Liu,S., Attix,C., Andreise,T., Frankheim,M.,
Amico-Reller,G., Coefield,J., Duarte,S., Lucas,S., Bruce,R.,
Thomas,P., Quan,G., Krommiller,B., Arellano,A., Sanders,C., Ow,D.,
Nolan,M., Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
TITLE Sequence analysis of a 1.9 Mb contig in 19p12 between UBA52 and
D19S455
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 45858)
AUTHORS Lamerdin,J.E.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-1998) Joint Genome Institute, Lawrence Livermore
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
REFERENCE 3 (bases 1 to 45858)
AUTHORS Lamerdin,J.E.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1998) Joint Genome Institute, Lawrence Livermore
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT Map and sequence are oriented from p telomere to centromere. Cosmid
R28177 overlaps cosmid R24069 (AC004400) to the left from bases 1
to 7,688 of this accession and overlaps cosmid R30064 (AC003107) to
the right from bases 33,879 to 45,858. Additional map and sequence
information may be obtained at:
http://www-bio.lnlnl.gov/bbrp/genome/genome.html.

FEATURES
Location/Qualifiers
1..45858
/organism="Homo sapiens"
/mol_type="genomic DNA"

```
/db_xref="taxon:9606"
/chromosome="19"
/map="19p12 between UBA52 and D19S455"
/clone="R28177"
/cell_line="5HL2-B"
/clone_lib="LL19NC03 R chromosome 19-specific cosmid
library"
/note="LL19NC03 cosmid library constructed at LLNL from
flow-sorted chromosomes from hybrid 5HL2-B, which carries
chromosome 19 as its only human chromosome."
1171..1328
/rpt_family="MLT1C"
repeat_region
1333..1636
/rpt_family="AluSg"
repeat_region
1652..1953
/rpt_family="AluJ"
repeat_region
1987..2287
/rpt_family="AluY"
repeat_region
2295..2396
/rpt_family="MLT1B"
repeat_region
complement(2407..2712)
/rpt_family="AluSx"
repeat_region
2729..2815
/rpt_family="MLT1B"
repeat_region
complement(3193..3502)
/rpt_family="AluSg"
repeat_region
complement(3503..3799)
/rpt_family="AluSx"
repeat_region
3831..4132
/rpt_family="AluY"
repeat_region
4150..4447
/rpt_family="AluJb"
repeat_region
complement(5635..5669)
/rpt_family="(GGGA)n"
repeat_region
complement(5670..5844)
/rpt_family="AluJb"
repeat_region
complement(5846..6252)
/rpt_family="MLT1C"
repeat_region
6843..7236
/rpt_family="MLT1A2"
repeat_region
7496..7788
/rpt_family="AluJ"
repeat_region
8443..8516
/rpt_family="L1"
repeat_region
8521..8800
/rpt_family="AluSx"
repeat_region
8802..9028
/rpt_family="L1MB6"
repeat_region
9105..9384
/rpt_family="AluJ"
repeat_region
12243..12311
/rpt_family="MLT1C"
repeat_region
12312..12725
/rpt_family="MER28"
repeat_region
12729..13023
/rpt_family="AluJb"
repeat_region
13053..13124
/rpt_family="TIGGER2"
repeat_region
13208..13506
/rpt_family="AluSg"
repeat_region
complement(14180..14232)
/rpt_family="(TGAA)n"
repeat_region
14533..14657
/rpt_family="MER77"
repeat_region
14658..14953
/rpt_family="AluSx"
repeat_region
14954..15050
/rpt_family="MER77"
repeat_region
15124..15356
/rpt_family="MER77"
repeat_region
15409..15709
/rpt_family="AluSx"
repeat_region
join(<17597..17713,18795..18842,20510..20647,21740..21801,
24468..24562,28187..28272,34238..34278,34695..34915,
40091..40215,43172..43480)
/product="Homo sapiens mRNA for KIAA0616 protein, partial
cds"
/note="non-consensus splice sites at positions 20647 and
21740"
Join(<17597..17713,18795..18842,20510..20647,21740..21801,
24468..24562,28187..28272,34238..34278,34695..34915,
40091..40215,43172..43480)
/note="non-consensus splice sites at positions 20647 and
21740"
/codon_start=1
/product="KIAA0616 protein"
/protein_id="AAC97072.1"
/db_xref="GI:4028930"
/translation="LQLKQSYLQLGPRGQYVGGSLPNVQIGSGTMDLPFQPSGFL
GEALAAPVSLTPFQSSGLDTSRTTRHHGLVDVYRERGLRGLSPHRLPSVDKRGROA
DSCPYGMVLSPPADTSWRTNSDSALHQSTMPTQEPSSQSGDVHOKRVLLITVP
QNETTSEADKNLSKQAWDKYKGRPKSCVEVGINIFPSAQDQNTTALIPATHTCG
SLPDLTNIHFPSPLTFLDPEEPTFPALSSSSSTGNLAANTHIGIGAGCGMSTPGS
SPQHRPAGVSLSLSTEARQASPTLSPLSPITQAVAMDALSLEQLPYAFPTQGS
QQPPPPQPPPPPPASQQPPPPPPQAPVRLPFGGGLPSASLTRGQPPPLAVTVP
SSLQSPENPGQPSMGIDIAS"
19161..19219
/rpt_family="(CA)n"
repeat_region
22042..22331
/rpt_family="AluSx"
repeat_region
22483..22781
/rpt_family="AluJ"
repeat_region
22820..22909
/rpt_family="L1MB3"
repeat_region
complement(22944..23205)
/rpt_family="AluSg"
repeat_region
23222..23405
/rpt_family="AluJ"
repeat_region
23407..23453
/rpt_family="(CA)n"
repeat_region
complement(23527..23704)
/rpt_family="MIR"
repeat_region
complement(24144..24250)
/rpt_family="FLAM_A"
repeat_region
25394..25421
/rpt_family="(CGA)n"
repeat_region
25929..26230
/rpt_family="AluSg"
repeat_region
26585..26854
/rpt_family="AluSx"
repeat_region
complement(27004..27305)
/rpt_family="AluJb"
repeat_region
complement(27326..27624)
/rpt_family="AluSx"
repeat_region
27671..27982
/rpt_family="AluY"
repeat_region
28505..28807
/rpt_family="AluSx"
repeat_region
28935..29214
/rpt_family="AluJ"
repeat_region
complement(29612..29669)
/rpt_family="(TAAAA)n"
repeat_region
complement(29671..29955)
/rpt_family="AluSx"
repeat_region
complement(29956..30166)
/rpt_family="AluJ"
repeat_region
complement(30288..30588)
/rpt_family="AluSx"
repeat_region
complement(30590..30658)
/rpt_family="AluJ"
repeat_region
30739..31041
/rpt_family="AluSg"
repeat_region
31052..31314
/rpt_family="AluJ"
repeat_region
32281..32413
/rpt_family="AluJ"
repeat_region
mRNA
```



```

CDS
complement(1219..1533)
/genes="SC01700"
/notes="SC130A.21c, putative membrane protein, len: 104 aa.
Contains possible hydrophobic membrane spanning regions"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="CAB46798.1"
/db_xref="GI:5441774"
/db_xref="SPTRMEL:Q98252"
/translation="MSDTPVKQNTAFYQGVASFAVAMATAIGIYQLQADAVRA
FLAIVLYLTSFTLAKIIRKQGVQVRSVDQARLILVEHDFPKPGGTGAG
RP"

gene
complement(1609..2760)
/genes="SC01701"
/notes="SC130A.22c, probable acyl-CoA dehydrogenase, len:
383 aa; similar to many both prokaryote and eukaryote eg.
SW:ACDB BACSU acyl-CoA dehydrogenase from Bacillus
subtilis (379 aa) fasta scores; opt: 109.1 z-score:
1197.8, E(): 0, (43.8% identity in 377 aa overlap) and
SW:ACDS RAT acyl-CoA dehydrogenase from Rattus norvegicus
(Rat) (412 aa) fasta scores; opt: 1079, z-score: 1184.2,
E(): 0, (44.8% identity in 372 aa overlap). Contains Pfam
match to entry PF00441 Acyl-CoA_dh, Acyl-CoA
dehydrogenase."
/codon_start=1
/transl_table=11
/product="putative acyl-CoA dehydrogenase"
/protein_id="CAB46799.1"
/db_xref="GI:5441775"
/db_xref="GOA:Q98251"
/db_xref="SPTRMEL:Q98251"
/translations="MNLSEIETQAVQLQARDFVEREIAPIHVVEWDRAEVDRSLVKK
LGVSLGLTIDQYSGSDHLYACLVTIELRGDSVVRGIVSLGVAKTIAAG
DEQKRWLPLGLTSYGVCGGLTEPGTSGDAGNLTRAVRDGDYVNGTWF.TNG
TWADVLLFARSDIACHQGVSAFLVPTDIPGLTRTHNGKLGRQATAEVLVDVR
VPASMLAPGKGFSAVAMALAKRMSVAGCVGIAQAALDAARVAGEREFGKTTA
HQVLQELISDIADLYDAALTRVADLIDRQPFVAVSSAKLFASEAARAANNA
LOVFGAYGIDYPAKLLRDARVMTLYEGTSIQIKLVICRALTGVSAP"
complement(1636..2748)
/genes="SC01701"
/notes="Pfam match to entry PF00441 Acyl-CoA dh, Acyl-CoA
dehydrogenase, score 585.20, E-value 4.3e-166."
2850..3491
/genes="SC01702"
/notes="synonym: SC130A.23"
2850..3491
/genes="SC01702"
/notes="SC130A.23, possible transcriptional regulator, len:
213 aa; similar to many e.g. TR:Q53901 (EMBL:M64683)
regulatory protein from Streptomyces coelicolor (259 aa)
fasta scores; opt: 227, z-score: 273.2, E(): 6.9e-08,
(32.3% identity in 217 aa overlap). Contains Pfam match to
entry PF00440 tetr, Bacterial regulatory proteins, tetr
family and Prosite match to PS01081 Bacterial regulatory
proteins, tetr family signature. Contains possible
helix-turn-helix motif (+5.17 SD) 31-52aa."
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="CAB46800.1"
/db_xref="GI:5441776"
/db_xref="GOA:Q98250"
/db_xref="SPTRMEL:Q98250"
/translations="MARPRKLLSTDRIVETARALVDAGLAASVTRRLAELGVSGP
SLYNHFKTEILEAVDSVSAQVDSMPEDGEMRTALHDWAVSYRTALRDHPNIVP
VLAHGPPRRPALHLADAVGVAMVRAGMPAAQTSICALMRVFMVMSALGSPAGFVD
DAGAYDPADYVPHLQAHLLAEQKEIKDERAFETGLTALLDGLAQQAQVAVQV"
2889..3029
/misc_feature
/genes="SC01702"

```

```

/notes="Pfam match to entry PF00440 tetr, Bacterial
regulatory proteins, tetr family, score 45.00, E-value
9.2e-11."
2925..3017
/genes="SC01702"
/notes="PS01081 Bacterial regulatory proteins, tetr family
signature."
3555..4268
/genes="SC01703"
/notes="synonym: SC130A.24"
3555..4268
/genes="SC01703"
/notes="SC130A.24, possible transcriptional regulator, len:
237aa; similar to many eg. SW:MERR STRLI probable mercury
resistance operon repressor from Streptomyces lividans
(125 aa) fasta scores; opt: 144, z-score: 183.3, E():
0.007, (38.2% identity in 89 aa overlap). Contains Pfam
match to entry PF01022 HTH 5, Bacterial regulatory
protein, arSR family. Contains possible helix-turn-helix
motif (+3.92 SD) 37-58aa."
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="CAB46801.1"
/db_xref="GI:5441777"
/db_xref="GOA:Q98249"
/db_xref="SPTRMEL:Q98249"

Alignment Scores:
Pred. No.: 2.02e+04 Length: 283100
Score: 97.50 Matches: 28
Percent Similarity: 33.01% Conservative: 6
Best Local Similarity: 27.18% Mismatches: 20
Query Match: 21.38% Indels: 49
DB: 1 Gaps: 6

US-09-972-032-2 (1-79) x SC0939110 (1-283100)
QY 3 GlyArgProArgAGG-----ValSerAlaGlyCysGlyPheAla 15
Db 169946 GGGCGGGTCCGCGAGAGCTGGGGCCCGCGTTCGGGAGCAGCGCGGGTGTTCGTCGCG 170005
QY 16 AspaLahistp-----19
Db 170006 CCTCGCACTGGCTCGCGAGGAGCGGTGGAGCGGTGACCCCGCGAGCGCGCG 170065
QY 20 -----ThrGlyLeu-----TrpThrGlyLeuGlyGluGly 29
Db 170066 CCGCGGACTGGGTACCGCGGTGAGCGGTGCTGCTGCTGGAGC-----GGCGCGGG 170119
QY 30 GlnGluGlyGlyIleGlyProGluGlyGlnAlaSerProThrProAspCysAlaSerArg 49
Db 170120 CGCGGGCGCGGTGGGGCCCTCGCGGGGTGGCGCCCGCGCGCGCGCG 170176
QY 50 TrpProArgSerAlaSer-----Arg 56
Db 170177 TGGCGGAGAGCGGTTCACGACCGCGGAGCTGATGATCACCAGTTCGAGTTCGCGGGCGC 170236
QY 57 TrpProTrp 59
Db 170237 TGGCCCTGG 170245

RESULT 15
AC113007/c
LOCUS
DEFINITION Mus musculus clone RP23-349H1.3, WORKING DRAFT SEQUENCE, 11
AC113007
Mus musculus clone RP23-349H1.3, WORKING DRAFT SEQUENCE, 11
unordered pieces.
AC113007
HTG: HTGS_PHASE1; HTGS DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 193012)
Birren,B., Nussbaum,C. and Lander,E.
Mus musculus, clone RP23-349H13
Unpublished
2 (bases 1 to 193012)
Birren,B., Linton,L., Nussbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Canarata,J., Campoliano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collamore,A., Cook,A.,
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galegan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
Lander,T., Lehoczy,J., Levine,R., Liu,G., Maclean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Ratta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (25-FEB-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 193012)
Birren,B., Nussbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhgalter,B., Canarata,J., Chang,J., Choepel,Y.,
Collamore,A., Cooke,P., Corum,B., DeArelano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galegan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafz,N., Hagopian,B., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Manning,J., Matthews,C.,
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,J., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 1, 2003 this sequence version replaced gi:21306978.
All repeats were identified using RepeatMasker:
Smit, A.P.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L22678
Center clone name: 349_H_13
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 189402 bases at least Q40
Consensus quality: 190906 bases at least Q30

Consensus quality: 191584 bases at least Q20
Insert size: 210000; agarose-fp
Insert size: 192012; sum-of-contigs
Quality coverage: 6.4 in Q20 bases; agarose-fp
Quality coverage: 7.0 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 56113: contig of 56113 bp in length
* 56213: gap of 100 bp
* 56214 56868: contig of 655 bp in length
* 56869 58967: contig of 1999 bp in length
* 58968 59067: gap of 100 bp
* 59068 62701: contig of 3634 bp in length
* 62702 62801: gap of 100 bp
* 62802 67586: contig of 4785 bp in length
* 67587 67866: gap of 100 bp
* 67867 72769: contig of 5083 bp in length
* 72770 72869: gap of 100 bp
* 72870 84598: contig of 11729 bp in length
* 84599 84698: gap of 100 bp
* 84699 107921: contig of 23223 bp in length
* 107922 108021: gap of 100 bp
* 108022 141625: contig of 33604 bp in length
* 141626 141725: gap of 100 bp
* 141726 174255: contig of 32530 bp in length
* 174256 174355: gap of 100 bp
* 174356 193012: contig of 18657 bp in length.
*
* Location/Qualifiers
* 1..193012
* /organism="Mus musculus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10090"
* /clone="RP23-349H13"
* /clone_lib="RPCI-23 Female Mouse BAC"
*
* 1..56113
* /note="assembly_fragment"
* clone_end:SP6
* vector_side:left
* 56214..56868
* /note="assembly_fragment"
* 56969..58967
* /note="assembly_fragment"
* 59068..62701
* /note="assembly_fragment"
* 62802..67586
* /note="assembly_fragment"
* 67687..72769
* /note="assembly_fragment"
* 72870..84598
* /note="assembly_fragment"
* 84699..107921
* /note="assembly_fragment"
* 108022..141625
* /note="assembly_fragment"
* 141726..174255
* /note="assembly_fragment"
* 174356..193012
* /note="assembly_fragment"
* clone_end:T7
* vector_side:right
*
* 1..59e+04 Length: 193012
* 97.00 Matches: 27

ORIGIN

Alignment Scores:
Pred. No.:
Score:

```

Percent Similarity: 35.11%      Conservative: 6
Best Local Similarity: 28.72%    Mismatches: 41
Query Match: 21.27%             Indels: 20
DB: 2                            Gaps: 3

US-09-972-032-2 (1-79) x AC113007 (1-193012)

Qy 5 ProArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThrGlyLeuTrpThr 24
Db 53982 CCGCGGCGAGTCGGGGCCCCCGGGCGGAGCCTGCACGGCCACGTGAGAGGTGAGTGGCGG 53923
Qy 25 -----GlyLeuGlyGluGlyGlnGlyGlyIle 34
Db 53922 GCGGGCGCGCATTTCTCGGCTGCGGGCGGCGCTGCGGGCGCCTCGGGTGGCGCTCGG 53863
Qy 35 GlyProGluGlyGlnAlaSer-----ProThrProAsp 45
Db 53862 GGGTCCCGCGGGCGGACCGCGATCGCGGCTGCGGCTACCGGCCCTGTCTCCCGCCTCCC 53803
Qy 46 CysAlaSerArgTrpProArgSerAlaSerArg--TipProTrpSerAlaGlyLeuThr 64
Db 53802 TCGGGCCCCCGTGGGGCGCGGACAGGACTAGTGGCCGGTGAGCCGGGCGCAGGTGC 53743
Qy 65 ValArgAspArgProGlnLeuGlyGluLeuCysMetGlyArg 78
Db 53742 GTGGCCGAGTGCTGTGAACCTCGGTGACCTCTGCCAGGGTCGA 53701

```

Search completed: July 10, 2004, 17:50:05
Job time : 3147 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: July 10, 2004, 14:37:10 ; Search time 538 Seconds
(without alignments)
8968.670 Million cell updates/sec

Title: US-09-972-032-1
Perfect score: 990
Sequence: 1 ggaattgtcttcgagccaa.....aaaatgagcgccgaagt 990

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3183909 seqs, 2436941669 residues

Word size : 0
Total number of hits satisfying chosen parameters: 6367818

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : Published Applications NA.*
- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
 - 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
 - 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
 - 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
 - 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
 - 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
 - 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
 - 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
 - 9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
 - 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
 - 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
 - 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
 - 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
 - 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
 - 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
 - 16: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
 - 17: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
 - 18: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
 - 19: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	990	100.0	990	13	US-09-972-032-1
2	80	8.1	404	10	Sequence 1, Appli
3	55	5.6	114615	13	Sequence 1098, Ap
4	54	5.5	174448	13	Sequence 676, App
5	51	5.2	149480	10	Sequence 148, App
6	51	5.2	149480	10	Sequence 284, App
7	51	5.2	149480	12	Sequence 285, App
8	50	5.1	215221	13	Sequence 232, App
9	49	4.9	738	13	Sequence 1360, Ap
10	49	4.9	738	16	Sequence 23812, A
11	49	4.9	827	13	Sequence 23812, A
12	49	4.9	827	13	Sequence 167845,
13	49	4.9	827	13	Sequence 167846,
14	49	4.9	827	16	Sequence 167845,

Sequence 167846,
Sequence 167847,
Sequence 32, Appl
Sequence 1463, Ap
Sequence 264705,
Sequence 264706,
Sequence 264705,
Sequence 264706,
Sequence 1762, Ap
Sequence 1795, Ap
Sequence 1451, Ap
Sequence 4, Appl
Sequence 46, Appl
Sequence 46, Appl
Sequence 43293, A
Sequence 43293, A
Sequence 17785, A
Sequence 17785, A
Sequence 46783, A
Sequence 46783, A
Sequence 73773, A
Sequence 73773, A
Sequence 313211,
Sequence 73773, A
Sequence 313211,
Sequence 313211,
Sequence 254288,
Sequence 254289,
Sequence 254288,
Sequence 254289,
Sequence 330, App
Sequence 127, App
Sequence 1356, App

ALIGNMENTS

RESULT 1
US-09-972-032-1
Sequence 1, Application US/09972032
Publication No. US20020086361A1
GENERAL INFORMATION:
APPLICANT: Case Western Reserve University
APPLICANT: Montano, Monica
APPLICANT: Sutton, Amelia
TITLE OF INVENTION: A Modulator of Antiestrogen Pharmacology
FILE REFERENCE: 27708/04003
CURRENT APPLICATION NUMBER: US/09/972,032
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: US 60/238,190
PRIOR FILING DATE: 2000-10-05
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent in version 3.1
SEQ ID NO 1
LENGTH: 990
TYPE: DNA
ORGANISM: Homo sapiens
US-09-972-032-1

Query Match	100.0%	Score 990;	DB 13;	Length 990;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 990;	Conservative	0;	Mismatches	0;
Indels	0;	Gaps	0;	
Qy	1	GGAATGTCTCGAGGCCAAGAAATCGGAGGATCTCTGACCTCGTATCGCCCG	60	
Db	1	GGAATGTCTCGAGGCCAAGAAATCGGAGGATCTCTGACCTCGTATCGCCCG	60	
Qy	61	CCTCAGCTTCCCAAAGTCTGGATTACAGGCATGAGCACCGCGCCCGCCCGACAC	120	
Db	61	CCTCAGCTTCCCAAAGTCTGGATTACAGGCATGAGCACCGCGCCCGCCCGACAC	120	
Qy	121	CTAGTTTAAAGGCCCTGCTGTGTGCTGCGCGCTGCCCGCTCCAGCTGCCAGTGTG	180	
Db	121	CTAGTTTAAAGGCCCTGCTGTGTGCTGCGCGCTGCCCGCTCCAGCTGCCAGTGTG	180	

```
QY 181 GCGGCTCAGTCCCGCTTCCCATGTGTGGAGACCCGCTCGGTAAAGCGCTGGATGTGG 240
Db 181 GCGGCTCAGTCCCGCTTCCCATGTGTGGAGACCCGCTCGGTAAAGCGCTGGATGTGG 240
QY 241 CTTGCGTGTGATGCAATTTGACCGGCTCTGCACTGGGCTAGGGAAGGCGAGGGCGG 300
Db 241 CTTGCGTGTGATGCAATTTGACCGGCTCTGCACTGGGCTAGGGAAGGCGAGGGCGG 300
QY 301 AATTGGGCCCCAGGGCCAGGCTTCCGCGACCCCGACTGCGGCTCCCGTGGCCCGCGAG 360
Db 301 AATTGGGCCCCAGGGCCAGGCTTCCGCGACCCCGACTGCGGCTCCCGTGGCCCGCGAG 360
QY 361 CGGCTCCCGGTGGCTTGGAGTGCAGGTCTTACGCTCCGAGATCGTCCGCACTGGGCGA 420
Db 361 CGGCTCCCGGTGGCTTGGAGTGCAGGTCTTACGCTCCGAGATCGTCCGCACTGGGCGA 420
QY 421 GCTGTGATGGGCGTGGCTTAAGCCGCTGTTTGGTTAGCATTTGGCCAGCGGACTTAAG 480
Db 421 GCTGTGATGGGCGTGGCTTAAGCCGCTGTTTGGTTAGCATTTGGCCAGCGGACTTAAG 480
QY 481 TGTGTCTCTGAAGAGCATGGAATTTAGTCTGAGGGTCTGGAAGAGTATCCCGCGCC 540
Db 481 TGTGTCTCTGAAGAGCATGGAATTTAGTCTGAGGGTCTGGAAGAGTATCCCGCGCC 540
QY 541 CACCATCAATGGCTTAGTCTAGGAGCGGCTGTGGGTGGGCGCTTAGGGCGAGGCG 600
Db 541 CACCATCAATGGCTTAGTCTAGGAGCGGCTGTGGGTGGGCGCTTAGGGCGAGGCG 600
QY 601 CAGACATACCCCAAGTGGTGGATTGTATPACCGCAAGGGCTTGGATCGAATCCCGCAAA 660
Db 601 CAGACATACCCCAAGTGGTGGATTGTATPACCGCAAGGGCTTGGATCGAATCCCGCAAA 660
QY 661 GACACTGGAGGCTGTGGCTCAGAGGGCGCGCAATCCAGTGTGTGGGCTTTAC 720
Db 661 GACACTGGAGGCTGTGGCTCAGAGGGCGCGCAATCCAGTGTGTGGGCTTTAC 720
QY 721 AGGAAAGAGCTCCACCTTCTGGAGTGTGCAGATGCGATCTAGTGTGTCCACCGATGG 780
Db 721 AGGAAAGAGCTCCACCTTCTGGAGTGTGCAGATGCGATCTAGTGTGTCCACCGATGG 780
QY 781 GAGCTGGGCGCGGAGATGCTGCCCGCAGTACAAAGCTGATTGACCTGGGCGCTCTG 840
Db 781 GAGCTGGGCGCGGAGATGCTGCCCGCAGTACAAAGCTGATTGACCTGGGCGCTCTG 840
QY 841 GACTTCCCTGATCTCTGCTTGCATCTCCAGCAAGTCTGCTCCCGTGGCTGCTTCAT 900
Db 841 GACTTCCCTGATCTCTGCTTGCATCTCCAGCAAGTCTGCTCCCGTGGCTGCTTCAT 900
QY 901 CCACCTCTCATTCTCTGCTTCCAGTAAATTCAGATCTGGGTGCAAAAAAAA 960
Db 901 CCACCTCTCATTCTCTGCTTCCAGTAAATTCAGATCTGGGTGCAAAAAAAA 960
QY 961 AAAAAAATAAATGAGCGGCGCAAGTT 990
Db 961 AAAAAAATAAATGAGCGGCGCAAGTT 990
```

RESULT 2

```
US-09-764-891-1098
; Sequence 1098, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1098
; LENGTH: 404
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (369)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (378)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (380)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-1098
```

```
Query Match 8.1%; Score 80; DB 10; Length 404;
Best Local Similarity 100.0%; Pred. No. 3.3e-30;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 872 CAAAGTCTGTCCCGTTGGCTTCCATCCACTCTCTCACTTCTGCTTCCAGATAA 931
Db 8 CAAAGTCTGTCCCGTTGGCTTCCATCCACTCTCTCACTTCTGCTTCCAGATAA 67
QY 932 AATTGCAAGATCTCTGTGTC 951
Db 68 AATTGCAAGATCTCTGTGTC 87
```

RESULT 3

```
US-10-087-192-676
; Sequence 676, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 676
; LENGTH: 114615
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1) ... (114615)
; OTHER INFORMATION: n = A,T,C or G
```

US-10-087-192-676

```
Query Match 5.6%; Score 55; DB 13; Length 114615;
Best Local Similarity 100.0%; Pred. No. 8.4e-18;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 38 CTCCTGACCTCGTGTATCCGCGCTCAGCTTCCCAAAGTGTGGGATTACAGGC 92
Db 57568 CTCCTGACCTCGTGTATCCGCGCTCAGCTTCCCAAAGTGTGGGATTACAGGC 57622
```

RESULT 4

```
US-10-087-192-148/c
; Sequence 148, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: CANCER
```



```
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 148
; LENGTH: 174448
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(174448)
; OTHER INFORMATION: n = A,T,C or G
US-10-087-192-148

Query Match          5.5%; Score 54; DB 13; Length 174448;
Best Local Similarity 100.0%; Pred. No. 2.6e-17; Indels 0; Gaps 0;
Matches 54; Conservative 0; Mismatches 0;

QY 49 GTGATCGCGCGCTCAGCTTCCCAAAGTCTGGATTACAGCATGAGCCACC 102
Db 80771 GTGATCGCGCGCTCAGCTTCCCAAAGTCTGGATTACAGCATGAGCCACC 80718

RESULT 5
US-09-873-367C-284
; Sequence 284, Application US/09873367C
; Publication No. US20030165839A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; APPLICANT: Soppet, Daniel
; APPLICANT: Endress, Gregory
; APPLICANT: Augustus, Meana
; APPLICANT: Ebner, Reinhard
; APPLICANT: Carter, Kenneth
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
; FILE REFERENCE: 689290-64
; CURRENT APPLICATION NUMBER: US/09/873,367C
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: U.S. 60/236,891
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: U.S. 60/236,842
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: U.S. 60/244,867
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: U.S. 60/245,084
; PRIOR FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 284
; LENGTH: 149480
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-873-367C-284

Query Match          5.2%; Score 51; DB 10; Length 149480;
Best Local Similarity 100.0%; Pred. No. 8.8e-16; Indels 0; Gaps 0;
Matches 51; Conservative 0; Mismatches 0;

QY 57 CCCGCTCAGCTTCCCAAAGTCTGGATTACAGCATGAGCCACCGGCC 107
Db 76386 CCCGCTCAGCTTCCCAAAGTCTGGATTACAGCATGAGCCACCGGCC 76436

RESULT 6
US-09-873-367C-285
; Sequence 285, Application US/09873367C
; Publication No. US20030165839A1
; GENERAL INFORMATION:
; APPLICANT: Young, Paul
; APPLICANT: Soppet, Daniel
; APPLICANT: Endress, Gregory
; APPLICANT: Augustus, Meana
; APPLICANT: Ebner, Reinhard
; APPLICANT: Carter, Kenneth
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
; FILE REFERENCE: 689290-64
; CURRENT APPLICATION NUMBER: US/09/873,367C
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: U.S. 60/236,891
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: U.S. 60/236,842
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: U.S. 60/244,867
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: U.S. 60/245,084
; PRIOR FILING DATE: 2000-11-01
; NUMBER OF SEQ ID NOS: 1067
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 284
; LENGTH: 149480
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-873-367C-285

Query Match          5.2%; Score 51; DB 10; Length 149480;
Best Local Similarity 100.0%; Pred. No. 8.8e-16; Indels 0; Gaps 0;
Matches 51; Conservative 0; Mismatches 0;

QY 57 CCCGCTCAGCTTCCCAAAGTCTGGATTACAGCATGAGCCACCGGCC 107
Db 76386 CCCGCTCAGCTTCCCAAAGTCTGGATTACAGCATGAGCCACCGGCC 76436

RESULT 7
US-09-968-007A-232
; Sequence 232, Application US/09968007A
; Publication No. US20040115825A1
; GENERAL INFORMATION:
; APPLICANT: Ebner, Reinhard
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using
; FILE REFERENCE: 689290-71
; CURRENT APPLICATION NUMBER: US/09/968,007A
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: US/60/237,172
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: US/60/237,173
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: US/60/237,278
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: US/60/237,294
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: US/60/237,295
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: US/60/237,316
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 1001
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 232
; LENGTH: 149480
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-968-007A-232

Query Match          5.2%; Score 51; DB 12; Length 149480;
Best Local Similarity 100.0%; Pred. No. 8.8e-16; Indels 0; Gaps 0;
Matches 51; Conservative 0; Mismatches 0;

QY 57 CCCGCTCAGCTTCCCAAAGTCTGGATTACAGCATGAGCCACCGGCC 107
Db 76386 CCCGCTCAGCTTCCCAAAGTCTGGATTACAGCATGAGCCACCGGCC 107
```

```
Db 76386 CCAGCTCTCCTGACCTCCCAAGTCTGGGATTACAGGCATGAGCCACCGCGGC 76436
|||||
RESULT 8
US-10-087-192-1360
; Sequence 1360, Application US/10087192
; Publication No. US20020182586A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1360
; LENGTH: 215221
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-087-192-1360

Query Match 5.1%; Score 50; DB 13; Length 215221;
Best Local Similarity 100.0%; Pred. No. 2.7e-15;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 CGATCTCTCCTGACCTCCGCGCCGCTCAGCTTCCCAAGTGTGG 83
|||||
Db 49208 CGATCTCTCCTGACCTCCGCGCCGCTCAGCTTCCCAAGTGTGG 49257
|||||

RESULT 9
US-10-027-632-23812
; Sequence 23812, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23812
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-23812

Query Match 4.9%; Score 49; DB 13; Length 738;
Best Local Similarity 100.0%; Pred. No. 1.8e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 CGATCTCTCCTGACCTCCGCGCCGCTCAGCTTCCCAAGTGTGG 82
|||||
Db 368 CGATCTCTCCTGACCTCCGCGCCGCTCAGCTTCCCAAGTGTGG 416
|||||

RESULT 11
US-10-027-632-167845/c
; Sequence 167845, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23812
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-23812

Query Match 4.9%; Score 49; DB 13; Length 738;
Best Local Similarity 100.0%; Pred. No. 1.8e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 34 CGATCTCTCCTGACCTCCGCGCCGCTCAGCTTCCCAAGTGTGG 82
|||||
Db 368 CGATCTCTCCTGACCTCCGCGCCGCTCAGCTTCCCAAGTGTGG 416
|||||

RESULT 10
US-10-027-632-23812
; Sequence 23812, Application US/10027632
; Publication No. US20020204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23812
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-23812

Query Match 4.9%; Score 49; DB 16; Length 738;
Best Local Similarity 100.0%; Pred. No. 1.8e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 CGATCTCTCCTGACCTCCGCGCCGCTCAGCTTCCCAAGTGTGG 82
|||||
Db 368 CGATCTCTCCTGACCTCCGCGCCGCTCAGCTTCCCAAGTGTGG 416
|||||

RESULT 11
US-10-027-632-167845/c
; Sequence 167845, Application US/10027632
; Publication No. US20020198371A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23812
; LENGTH: 738
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-23812

Query Match 4.9%; Score 49; DB 16; Length 738;
Best Local Similarity 100.0%; Pred. No. 1.8e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167845
; LENGTH: 827
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-167845

Query Match
Best Local Similarity 4.9%; Score 49; DB 13; Length 827;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 GCCCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 104
|||||
Db 704 GCCCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 656

RESULT 12
US-10-027-632-167846/c
; Sequence 167846, Application US/10027632
; Publication No. US2002019837A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167846
; LENGTH: 827
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-167846

Query Match
Best Local Similarity 4.9%; Score 49; DB 13; Length 827;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 GCCCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 104
|||||
Db 704 GCCCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 656

RESULT 13
US-10-027-632-167847/c
; Sequence 167847, Application US/10027632
; Publication No. US2002019837A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
```

```
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167847
; LENGTH: 827
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-167847

Query Match
Best Local Similarity 4.9%; Score 49; DB 13; Length 827;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 GCCCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 104
|||||
Db 704 GCCCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 656

RESULT 14
US-10-027-632-167845/c
; Sequence 167845, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167845
; LENGTH: 827
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-167845

Query Match
Best Local Similarity 4.9%; Score 49; DB 16; Length 827;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 GCCCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 104
|||||
Db 704 GCCCGCCTCAGCTTCCCAAAGTGTGGATTACAGGCATGAGCCACCGC 656

RESULT 15
US-10-027-632-167846/c
; Sequence 167846, Application US/10027632
```

```
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167846
; LENGTH: 827
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-167846

Query Match      4.9%; Score 49; DB 16; Length 827;
Best Local Similarity 100.0%; Pred. No. 1.8e-14;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      56 GCCGCGCTCAGCTCCCAAAGTGTGGGATTACAGGCATGAGCCACGC 104
      |||||||
Db      704 GCCGCGCTCAGCTCCCAAAGTGTGGGATTACAGGCATGAGCCACGC 656
```

Search completed: July 10, 2004, 16:51:03
Job time : 544 secs

29: em_vt:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	79	100.0	195476	0	9	AC135044
2	79	100.0	195476	2	AC135044	AC135044
3	11	13.9	195476	1	AC135044	AC135044
C 4	9	11.4	195476	1	AC135044	AC135044
C 5	9	11.4	195476	6	AC135044	AC135044
6	9	11.4	195476	3	AC135044	AC135044
7	9	11.4	195476	3	AC135044	AC135044
8	9	11.4	195476	2	AC135044	AC135044
9	9	11.4	195476	2	AC135044	AC135044
10	9	11.4	195476	2	AC135044	AC135044
11	9	11.4	195476	9	AC135044	AC135044
C 12	9	11.4	195476	2	AC135044	AC135044
C 13	9	11.4	195476	2	AC135044	AC135044
C 14	9	11.4	195476	9	AC135044	AC135044
15	9	11.4	195476	2	AC135044	AC135044
16	9	11.4	195476	2	AC135044	AC135044
17	9	11.4	195476	2	AC135044	AC135044
18	9	11.4	195476	2	AC135044	AC135044
19	9	11.4	195476	9	AC135044	AC135044
C 20	9	11.4	195476	10	AC135044	AC135044
C 21	9	11.4	195280	9	AC073869	AC073869
C 22	9	11.4	195280	2	AC073869	AC073869
C 23	9	11.4	195280	2	AC073869	AC073869
C 24	9	11.4	195280	1	AC073869	AC073869
25	8	10.1	195280	9	AC073869	AC073869
26	8	10.1	195280	9	AC073869	AC073869
C 27	8	10.1	195280	6	AC073869	AC073869
28	8	10.1	195280	1	AC073869	AC073869
C 29	8	10.1	195280	6	AC073869	AC073869
C 30	8	10.1	195280	6	AC073869	AC073869
31	8	10.1	195280	10	AC073869	AC073869
C 32	8	10.1	195280	9	AC073869	AC073869
33	8	10.1	195280	4	AC073869	AC073869
C 34	8	10.1	195280	9	AC073869	AC073869
35	8	10.1	195280	9	AC073869	AC073869
C 36	8	10.1	195280	6	AC073869	AC073869
37	8	10.1	195280	9	AC073869	AC073869
38	8	10.1	195280	9	AC073869	AC073869
39	8	10.1	195280	9	AC073869	AC073869
40	8	10.1	195280	9	AC073869	AC073869
41	8	10.1	195280	9	AC073869	AC073869
42	8	10.1	195280	9	AC073869	AC073869
43	8	10.1	195280	9	AC073869	AC073869
44	8	10.1	195280	9	AC073869	AC073869
45	8	10.1	195280	9	AC073869	AC073869

ALIGNMENTS

RESULT 1

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model
Run on: July 10, 2004, 18:59:03 ; Search time 2822 Seconds
(without alignments)
1213.359 Million cell updates/sec

Title: US-09-972-032-2
Perfect score: 79
Sequence: 1 MGRPRRVSAGGFADAHWT.....SAGLTVRDRPOLGELCMGRG 79

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Word size: 1
Total number of hits satisfying chosen parameters: 6933665

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09972032/runat_06072004_121534_8872/app_query.fasta_1.263
-DB=GenEmbl -OPMT=fastap -SUFFIX=rge -MINMATCH=0_1 -LOOPECL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09972032@cgn_1_1_3731 @runat_06072004_121534_8872 -NCFU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb_ba.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*

Consensus quality: 188095 bases at least Q20
 Estimated insert size: 170000; agarose-fp estimation
 Estimated insert size: 193376; sum-of-contigs estimation
 Quality coverage: 7.63 in Q20 bases; agarose-fp estimation
 Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation.

* NOTE: This is a 'working draft' sequence. It currently
 consists of 22 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 be preserved.

1 1202: contig of 1202 bp in length
 * 1203 1302: gap of unknown length
 * 1303 2468: contig of 1166 bp in length
 * 2469 2568: gap of unknown

Alignment Scores:

Pred. No.: 45.3 Length: 195476
 Score: 9.00 Matches: 5
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 11.39% Indels: 0
 DB: 6 Gaps: 0

US-09-972-032-2 (1-79) x AC135044 (1-195476)

Qy 32 GlyGlyIleGlyPro 36

Db 162005 GCGGGTATTGGCGG 161991

RESULT 6

AC135044

LOCUS

DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT

SEQUENCE, 22 unordered pieces.

AC135044

VERSION AC135044.1 GI:23505535

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

DOE Joint Genome Institute.

Sequencing of Human Chromosome 16

Unpublished

2 (bases 1 to 195476)

DOE Joint Genome Institute.

Direct Submission

Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: <http://www.jgi.doe.gov>

Project Information

Center Project Name: 809609

Center clone name: CITB-El_2551B20

Summary Statistics

Consensus quality: 171229 bases at least Q40

Consensus quality: 182638 bases at least Q30

Consensus quality: 188095 bases at least Q20

Estimated insert size: 170000; agarose-fp estimation

Estimated insert size: 193376; sum-of-contigs estimation

Quality coverage: 7.63 in Q20 bases; agarose-fp estimation

Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation.

* NOTE: This is a 'working draft' sequence. It currently

consists of 22 contigs. The true order of the pieces

is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 be preserved.

1 1202: contig of 1202 bp in length
 * 1203 1302: gap of unknown length
 * 1303 2468: contig of 1166 bp in length
 * 2469 2568: gap of unknown

Alignment Scores:

Pred. No.: 198 Length: 195476
 Score: 9.00 Matches: 5
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 11.39% Indels: 0
 DB: 3 Gaps: 0

US-09-972-032-2 (1-79) x AC135044 (1-195476)

Qy 32 GlyGlyIleGlyPro 36

Db 98566 GGTGGATAGGCCCC 98580

RESULT 7

AC135044

LOCUS

DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT

SEQUENCE, 22 unordered pieces.

AC135044

VERSION AC135044.1 GI:23505535

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

DOE Joint Genome Institute.

Sequencing of Human Chromosome 16

Unpublished

2 (bases 1 to 195476)

DOE Joint Genome Institute.

Direct Submission

Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: <http://www.jgi.doe.gov>

Project Information

Center Project Name: 809609

Center clone name: CITB-El_2551B20

Summary Statistics

Consensus quality: 171229 bases at least Q40

Consensus quality: 182638 bases at least Q30

Consensus quality: 188095 bases at least Q20

Estimated insert size: 170000; agarose-fp estimation

Estimated insert size: 193376; sum-of-contigs estimation

Quality coverage: 7.63 in Q20 bases; agarose-fp estimation

Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation.

* NOTE: This is a 'working draft' sequence. It currently

consists of 22 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence.

* as soon as it is available and the accession number will

be preserved.

1 1202: contig of 1202 bp in length

* 1203 1302: gap of unknown length

* 1303 2468: contig of 1166 bp in length


```

*      2469      2568: gap of unknown

Alignment Scores:
Pred. No.:      206      Length:      195476
Score:          9.00      Matches:      5
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      11.39%      Indels: 0
DB:              3      Gaps: 0

US-09-972-032-2 (1-79) x AC135044 (1-195476)
LOCUS      AC135044
DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT
ACCESSION      AC135044
VERSION      AC135044.1 GI:23505535
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 195476)
AUTHORS      DOE Joint Genome Institute.
TITLE      Sequencing of Human Chromosome 16
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 195476)
AUTHORS      DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL      Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 809609
Center clone name: CITB-EL_2551B20
-----
Summary Statistics
Consensus quality: 171229 bases at least Q40
Consensus quality: 182638 bases at least Q30
Consensus quality: 188095 bases at least Q20
Estimated insert size: 170000; agarose-fp estimation
Quality coverage: 7.63 in Q20 bases; sum-of-contigs estimation
Quality coverage: 6.7 in Q20 bases; agarose-fp estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*      1      1202: contig of 1202 bp in length
*      1203      1302: gap of unknown length
*      1303      2468: contig of 1166 bp in length
*      2469      2568: gap of unknown

Alignment Scores:
Pred. No.:      1.57e+03      Length:      195476
Score:          9.00      Matches:      5
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      11.39%      Indels: 0
DB:              2      Gaps: 0

US-09-972-032-2 (1-79) x AC135044 (1-195476)
LOCUS      AC135044
DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT
ACCESSION      AC135044
VERSION      AC135044.1 GI:23505535
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 195476)
AUTHORS      DOE Joint Genome Institute.
TITLE      Sequencing of Human Chromosome 16
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 195476)
AUTHORS      DOE Joint Genome Institute.
TITLE      Direct Submission
JOURNAL      Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 809609
Center clone name: CITB-EL_2551B20
-----
Summary Statistics
Consensus quality: 171229 bases at least Q40
Consensus quality: 182638 bases at least Q30
Consensus quality: 188095 bases at least Q20
Estimated insert size: 170000; agarose-fp estimation
Quality coverage: 7.63 in Q20 bases; sum-of-contigs estimation
Quality coverage: 6.7 in Q20 bases; agarose-fp estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*      1      1202: contig of 1202 bp in length
*      1203      1302: gap of unknown length
*      1303      2468: contig of 1166 bp in length
*      2469      2568: gap of unknown

Alignment Scores:
Pred. No.:      1.04e+03      Length:      195476
Score:          9.00      Matches:      5
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      11.39%      Indels: 0
DB:              0

```

```

RESULT 10
AC135044
LOCUS
DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT
AC135044
ACCESSION AC135044.1 GI:23505535
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 195476)
DIRECT SUBMISSION
DOE Joint Genome Institute.
JOURNAL
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 195476)
DIRECT SUBMISSION
DOE Joint Genome Institute.
JOURNAL
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 195476)
DIRECT SUBMISSION
DOE Joint Genome Institute.
JOURNAL
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 809609
Center clone name: C1TB-El_2551B20
-----
Summary Statistics
Consensus quality: 171229 bases at least Q40
Consensus quality: 182638 bases at least Q30
Consensus quality: 188095 bases at least Q20
Estimated insert size: 170000; agarose-fp estimation
Quality coverage: 7.63 in Q20 bases; agarose-fp estimation
Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1202: contig of 1202 bp in length
* 1203 1302: gap of unknown length
* 1303 2468: contig of 1166 bp in length
* 2469 2568: gap of unknown
Alignment Scores:
Pred. No.: 1.57e+03 Length: 195476
Score: 9.00 Matches: 5
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.39% Indels: 0
DB: 2 Gaps: 0
US-09-972-032-2 (1-79) x AC135044 (1-195476)
QY 32 GlyGlytGlyPro 36
Db 98566 GGTGGATAGGCCCC 98580
RESULT 11
AC135044
LOCUS
DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT
AC135044
ACCESSION AC135044.1 GI:23505535
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 195476)
DIRECT SUBMISSION
DOE Joint Genome Institute.
JOURNAL
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 809609
Center clone name: C1TB-El_2551B20
-----
Summary Statistics
Consensus quality: 171229 bases at least Q40
Consensus quality: 182638 bases at least Q30
Consensus quality: 188095 bases at least Q20
Estimated insert size: 170000; agarose-fp estimation
Quality coverage: 7.63 in Q20 bases; agarose-fp estimation
Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1202: contig of 1202 bp in length
* 1203 1302: gap of unknown length
* 1303 2468: contig of 1166 bp in length
* 2469 2568: gap of unknown
Alignment Scores:
Pred. No.: 1.57e+03 Length: 195476
Score: 9.00 Matches: 5
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.39% Indels: 0
DB: 2 Gaps: 0
US-09-972-032-2 (1-79) x AC135044 (1-195476)
QY 32 GlyGlytGlyPro 36
Db 98566 GGTGGATAGGCCCC 98580

```

```

SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 195476)
DIRECT SUBMISSION
DOE Joint Genome Institute.
JOURNAL
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 195476)
DIRECT SUBMISSION
DOE Joint Genome Institute.
JOURNAL
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 809609
Center clone name: C1TB-El_2551B20
-----
Summary Statistics
Consensus quality: 171229 bases at least Q40
Consensus quality: 182638 bases at least Q30
Consensus quality: 188095 bases at least Q20
Estimated insert size: 170000; agarose-fp estimation
Quality coverage: 7.63 in Q20 bases; agarose-fp estimation
Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1202: contig of 1202 bp in length
* 1203 1302: gap of unknown length
* 1303 2468: contig of 1166 bp in length
* 2469 2568: gap of unknown
Alignment Scores:
Pred. No.: 1.78e+03 Length: 195476
Score: 9.00 Matches: 5
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.39% Indels: 0
DB: 9 Gaps: 0
US-09-972-032-2 (1-79) x AC135044 (1-195476)
QY 32 GlyGlytGlyPro 36
Db 98566 GGTGGATAGGCCCC 98580
RESULT 12
AC135044/c
LOCUS
DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT
AC135044
ACCESSION AC135044.1 GI:23505535
VERSION
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE
ORGANISM Homo sapiens (human)
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 195476)
DIRECT SUBMISSION
DOE Joint Genome Institute.
JOURNAL
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16

```

REFERENCE 2 (bases 1 to 195476)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI
 Web site: http://www.jgi.doe.gov

 Project Information
 Project Name: 809609
 Center clone name: CITB-EL_2551B20

 Summary Statistics
 Consensus quality: 171229 bases at least Q40
 Consensus quality: 182638 bases at least Q30
 Consensus quality: 188095 bases at least Q20
 Estimated insert size: 170000; agarose-fp estimation
 Estimated insert size: 193376; sum-of-contigs estimation
 Quality coverage: 6.73 in Q20 bases; agarose-fp estimation
 Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 22 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be-preserved.
 * 1 1202: contig of 1202 bp in length
 * 1203 1302: gap of unknown length
 * 1303 2468: contig of 1166 bp in length
 * 2469 2568: gap of unknown

Alignment Scores:
 Pred. No.: 1.97e+03 Length: 195476
 Score: 9.00 Matches: 5
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 11.39% Indels: 0
 DB: 2 Gaps: 0

US-09-972-032-2 (1-79) x AC135044 (1-195476)
 QY 32 GlyGlyIleGlyPro 36
 Db 162005 GGGGGTATGGCCG 161991

RESULT 13
 LOCUS AC135044/c
 DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT
 AC135044 195476 bp DNA linear HTG 05-OCT-2002
 Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT
 SEQUENCE, 22 unordered pieces.
 AC135044
 AC135044.1 GI:23505535
 HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEPIN.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 195476)
 DOE Joint Genome Institute.
 Sequencing of Human Chromosome 16
 Unpublished
 REFERENCE 2 (bases 1 to 195476)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT -----Genome Center
 Center: Joint Genome Institute
 Center Code: JGI

Consensus quality: 182638 bases at least Q30
 Consensus quality: 188095 bases at least Q20
 Estimated insert size: 170000; agarose-fp estimation
 Estimated insert size: 193376; sum-of-contigs estimation
 Quality coverage: 7.63 in Q20 bases; agarose-fp estimation
 Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation.
 * NOTE: This is a 'working draft' sequence. It currently

* consists of 22 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1202: contig of 1202 bp in length
 * 1203 1302: gap of unknown length
 * 1303 2468: contig of 1166 bp in length
 * 2469 2568: gap of unknown

Alignment Scores:

Pred. No.:	2.18e+03	Length:	195476
Score:	9.00	Matches:	5
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	11.39%	Indels:	0
DB:	9	Gaps:	0

US-09-972-032-2 (1-79) x AC135044 (1-195476)

QY 32 GlyGlyIleGlyPro 36

Db 162005 GGGGTATTGGCGCG 161991

RESULT 15

AC135044

LOCUS

AC135044 195476 bp DNA linear HTG 05-OCT-2002
 DEFINITION Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT
 SEQUENCE, 22 unordered pieces.

AC135044

AC135044.1 GI:23505535

HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

SOURCE Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 195476)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 16

Unpublished

REFERENCE 2 (bases 1 to 195476)

DOE Joint Genome Institute.

Direct Submission

Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint

Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

-----Genome Center

Center: Joint Genome Institute

Center Code: JGI

Web site: <http://www.jgi.doe.gov>

Project Information

Center Project Name: 809609

Center clone name: CITB-B1.2551B20

Summary Statistics

Consensus quality: 171229 bases at least Q40

Consensus quality: 182638 bases at least Q30

Consensus quality: 188095 bases at least Q20

Estimated insert size: 170000; agarose-fp estimation

Estimated insert size: 193376; sum-of-contigs estimation

Quality coverage: 7.63 in Q20 bases; agarose-fp estimation

Quality coverage: 6.7 in Q20 bases; sum-of-contigs estimation.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 22 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 1202: contig of 1202 bp in length
 * 1203 1302: gap of unknown length
 * 1303 2468: contig of 1166 bp in length
 * 2469 2568: gap of unknown

Alignment Scores:

Pred. No.:	2.27e+03	Length:	195476
Score:	9.00	Matches:	5
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	11.39%	Indels:	0
DB:	2	Gaps:	0

US-09-972-032-2 (1-79) x AC135044 (1-195476)

QY 32 GlyGlyIleGlyPro 36

Db 98566 GGTGGATAGCCCC 98580

Search completed: July 10, 2004, 20:57:09

Job time : 3260 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 12, 2004, 19:22:20 / Search time 399 Seconds
(without alignments)
965.004 Million cell updates/sec

Title: US-09-972-032-2

Perfect score: 79

Sequence: 1 MGRPRVYAGCGFADAHWT.....SAGLTVRDRLQGLCMGRG 79

Scoring table:

OLIGO	
Xgapex 60.0	Xgapex 60.0
Ygapex 60.0	Ygapex 60.0
Fgapex 6.0	Fgapex 7.0
Delop 6.0	Delox 7.0

Searched: 3183909 seqs, 2436941669 residues

Word size: 1

Total number of hits satisfying chosen parameters: 6358483

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

```

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09972032/runat_06072004_121537_8984/app_query.fasta_1.263
-DB=Published Applications NA -QFMT=fastcap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR_MIN=1
-ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAFSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09972032@cgn_1_1_511@runat_06072004_121537_8984
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=6 -DELOP=6 -DELEXT=7

```

Database :

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*

18: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

19: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

ALIGNMENTS

RESULT 1

US-09-972-032-1

Sequence 1, Application US/09972032

Publication No. US20020086361A1

GENERAL INFORMATION:

APPLICANT: Case Western Reserve University

APPLICANT: Montano, Monica

APPLICANT: Sutton, Amelia

TITLE OF INVENTION: A Modulator of Antiestrogen Pharmacology

FILE REFERENCE: 27708/04003

CURRENT APPLICATION NUMBER: US/09/972,032

CURRENT FILING DATE: 2001-10-05

PRIOR APPLICATION NUMBER: US 60/238,190

PRIOR FILING DATE: 2000-10-05

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn version 3.1

SEQ ID NO 1

LENGTH: 990

TYPE: DNA

ORGANISM: Homo sapiens

US-09-972-032-1

Alignment Scores:

```
Pred. No.: 1.68e-69 Length: 990
Score: 79.00 Matches: 79
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x US-09-972-032-1 (1-990)

Qy 1 MetCysGlyArgProArgArgValSerAlaGlyCysGlyPheAlaaspAlahisTrpThr 20
Db 203 ATGTGTGGGACCGGTCGCGTAAGCGCTGGATGTGGCTTCGCTGATGCACATTGGACC 262
Qy 21 GlyLeuTrpThrGlyLeuGlyGlnGlnGlyGlyGlnGlyGlyGlnGlnAla 40
Db 263 GGGCTCTGGACTGGCTAGGGGAAGGCGAGGAGGGCGGAATTGGGCCCCGAGGGCCAGGCC 322
Qy 41 SerProThrProAspCysAlaSerArgTrpProArgSerAlaSerArgTrpProTpsr 60
Db 323 TCGCCGACCCCGACTGGCGCTCCCGTGGCGCCCGAGCGCTCCCGGTGGCCCTGGAGT 382
Qy 61 AlaGlyLeuThrValArgAspArgProGlnLeuGlyGlnLeuCysMetGlyArgGly 79
Db 383 CGAGGTCTTACCGTCCGAGATCGTCCGCAACTGGCGAGCTGTGCATGGGGGTGGC 439

RESULT 2
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown.
US-10-156-761-1

Alignment Scores:
Pred. No.: 81.4 Length: 9025608
Score: 11.00 Matches: 11
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 13.92% Indels: 0
DB: 15 Gaps: 0

US-09-972-032-2 (1-79) x US-10-156-761-1 (1-9025608)

Qy 48 SerArgTrpProArgSerAlaSerArgTrpPro 58
Db 4452379 TCGCGCTGGCCGAGAGTGCATCGGTGGCCA 4452411

RESULT 3
US-10-437-963-31286
; Sequence 31286, Application US/10437963
; Publication No. US20040123339A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 31286
; LENGTH: 250
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_35608C.1
US-10-437-963-31286

Alignment Scores:
Pred. No.: 25.2 Length: 250
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 17 Gaps: 0

US-09-972-032-2 (1-79) x US-10-437-963-31286 (1-250)

Qy 28 GluGlyGlnGlyGlyGlyGlyGlyGly 35
Db 130 GAAGGCAAGAGGAGGAGGATCGGT 153

RESULT 4
US-09-922-293-1599/c
; Sequence 1599, Application US/09922293
; Publication No. US20040123339A1
; GENERAL INFORMATION:
; APPLICANT: Comer, Timothy W.
; APPLICANT: Heck, Gregory R.
; APPLICANT: Liu, Jingsong
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Transcription in Plants
; FILE REFERENCE: 16517.254
; CURRENT APPLICATION NUMBER: US/09/922,293
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/067,000
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: US 60/069,472
; PRIOR FILING DATE: 1997-12-09
; PRIOR APPLICATION NUMBER: US 60/071,479
; PRIOR FILING DATE: 1998-01-13
; PRIOR APPLICATION NUMBER: US 60/074,201
; PRIOR FILING DATE: 1998-02-10
; PRIOR APPLICATION NUMBER: US 60/074,282
; PRIOR FILING DATE: 1998-02-10
; PRIOR APPLICATION NUMBER: US 60/074,280
; PRIOR FILING DATE: 1998-02-10
; PRIOR APPLICATION NUMBER: US 60/074,281
; PRIOR FILING DATE: 1998-02-10
; PRIOR APPLICATION NUMBER: US 60/074,566
; PRIOR FILING DATE: 1998-02-12
; PRIOR APPLICATION NUMBER: US 60/074,567
; PRIOR FILING DATE: 1998-02-12
; PRIOR APPLICATION NUMBER: US 60/074,565
; PRIOR FILING DATE: 1998-02-12
; PRIOR APPLICATION NUMBER: US 60/075,462
; PRIOR FILING DATE: 1998-02-19
```

;; PRIOR APPLICATION NUMBER: US 60/074,789
;; PRIOR FILING DATE: 1998-02-19
;; PRIOR APPLICATION NUMBER: US 60/075,459
;; PRIOR FILING DATE: 1998-02-19
;; PRIOR APPLICATION NUMBER: US 60/075,461
;; PRIOR FILING DATE: 1998-02-19
;; PRIOR APPLICATION NUMBER: US 60/075,464
;; PRIOR FILING DATE: 1998-02-19
;; PRIOR APPLICATION NUMBER: US 60/075,460
;; PRIOR FILING DATE: 1998-02-19
;; PRIOR APPLICATION NUMBER: US 60/075,463
;; PRIOR FILING DATE: 1998-02-19
;; PRIOR APPLICATION NUMBER: US 60/077,231
;; PRIOR FILING DATE: 1998-03-09
;; PRIOR APPLICATION NUMBER: US 60/077,229
;; PRIOR FILING DATE: 1998-03-09
;; PRIOR APPLICATION NUMBER: US 60/077,230
;; PRIOR FILING DATE: 1998-03-09
;; PRIOR APPLICATION NUMBER: US 60/078,368
;; PRIOR FILING DATE: 1998-03-18
;; PRIOR APPLICATION NUMBER: US 60/080,844
;; PRIOR FILING DATE: 1998-04-07
;; PRIOR APPLICATION NUMBER: US 60/083,067
;; PRIOR FILING DATE: 1998-04-27
;; PRIOR APPLICATION NUMBER: US 60/083,386
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: US 60/083,387
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: US 60/083,388
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: US 60/083,389
;; PRIOR FILING DATE: 1998-04-29
;; PRIOR APPLICATION NUMBER: US 60/085,224
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: US 60/085,223
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: US 60/085,222
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: US 60/085,533
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: US 60/086,186
;; PRIOR FILING DATE: 1998-05-21
;; PRIOR APPLICATION NUMBER: US 60/086,187
;; PRIOR FILING DATE: 1998-05-21
;; PRIOR APPLICATION NUMBER: US 60/086,185
;; PRIOR FILING DATE: 1998-05-21
;; PRIOR APPLICATION NUMBER: US 60/086,184
;; PRIOR FILING DATE: 1998-05-21
;; PRIOR APPLICATION NUMBER: US 60/086,183
;; PRIOR FILING DATE: 1998-05-21
;; PRIOR APPLICATION NUMBER: US 60/086,188
;; PRIOR FILING DATE: 1998-05-21
;; PRIOR APPLICATION NUMBER: US 60/089,524
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: US 60/089,810
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: US 60/089,814
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: US 60/089,808
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: US 60/089,812
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: US 60/089,807
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: US 60/089,806
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: US 60/089,813
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: US 60/089,811
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: US 60/089,793
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: US 60/091,405

;; PRIOR FILING DATE: 1998-06-30
;; PRIOR APPLICATION NUMBER: US 60/091,247
;; PRIOR FILING DATE: 1998-06-30
;; PRIOR APPLICATION NUMBER: US 60/099,667
;; PRIOR FILING DATE: 1998-09-09
;; PRIOR APPLICATION NUMBER: US 60/099,668
;; PRIOR FILING DATE: 1998-09-09
;; PRIOR APPLICATION NUMBER: US 60/099,670
;; PRIOR FILING DATE: 1998-09-09
;; PRIOR APPLICATION NUMBER: US 60/099,697
;; PRIOR FILING DATE: 1998-09-09
;; PRIOR APPLICATION NUMBER: US 60/100,674
;; PRIOR FILING DATE: 1998-09-16
;; PRIOR APPLICATION NUMBER: US 60/100,673
;; PRIOR FILING DATE: 1998-09-16
;; PRIOR APPLICATION NUMBER: US 60/100,672
;; PRIOR FILING DATE: 1998-09-16
;; PRIOR APPLICATION NUMBER: US 60/100,963
;; PRIOR FILING DATE: 1998-09-17
;; PRIOR APPLICATION NUMBER: US 60/101,131
;; PRIOR FILING DATE: 1998-09-21
;; PRIOR APPLICATION NUMBER: US 60/101,132
;; PRIOR FILING DATE: 1998-09-21
;; PRIOR APPLICATION NUMBER: US 60/101,130
;; PRIOR FILING DATE: 1998-09-21
;; PRIOR APPLICATION NUMBER: US 60/101,508
;; PRIOR FILING DATE: 1998-09-22
;; PRIOR APPLICATION NUMBER: US 60/101,344
;; PRIOR FILING DATE: 1998-09-22
;; PRIOR APPLICATION NUMBER: US 60/101,347
;; PRIOR FILING DATE: 1998-09-22
;; PRIOR APPLICATION NUMBER: US 60/101,343
;; PRIOR FILING DATE: 1998-09-22
;; PRIOR APPLICATION NUMBER: US 60/101,707
;; PRIOR FILING DATE: 1998-09-25
;; PRIOR APPLICATION NUMBER: US 60/104,126
;; PRIOR FILING DATE: 1998-10-13
;; PRIOR APPLICATION NUMBER: US 60/104,128
;; PRIOR FILING DATE: 1998-10-13
;; PRIOR APPLICATION NUMBER: US 60/104,127
;; PRIOR FILING DATE: 1998-10-13
;; PRIOR APPLICATION NUMBER: US 60/104,124
;; PRIOR FILING DATE: 1998-10-13
;; PRIOR APPLICATION NUMBER: US 60/109,018
;; PRIOR FILING DATE: 1998-11-18
;; PRIOR APPLICATION NUMBER: US 60/108,996
;; PRIOR FILING DATE: 1998-11-18
;; PRIOR APPLICATION NUMBER: US 09/199,129
;; PRIOR FILING DATE: 1998-11-24
;; PRIOR APPLICATION NUMBER: US 09/210,297
;; PRIOR FILING DATE: 1998-12-08
;; PRIOR APPLICATION NUMBER: US 60/111,981
;; PRIOR FILING DATE: 1998-12-11
;; PRIOR APPLICATION NUMBER: US 60/113,224
;; PRIOR FILING DATE: 1998-12-22
;; PRIOR APPLICATION NUMBER: US 09/229,413
;; PRIOR FILING DATE: 1999-01-12
;; NUMBER OF SEQ ID NOS: 3853
;; SEQ ID NO 1599
;; LENGTH: 254
;; TYPE: DNA
;; ORGANISM: Zea mays
US-09-922-293-1599

Alignment Scores:
Pred. No.: 25.5 Length: 254
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 12 Gaps: 0

US-09-972-032-2 (1-79) x US-09-922-293-1599 (1-254)

Qy 52 ArgSerAlaSerArgTrpProTyr 59
Db 31 CGATCAGCGAGCCGTTGCCCATGG 8

RESULT 5

US-10-156-761-6581
; Sequence 6581, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6581
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(507)
US-10-156-761-6581

Alignment Scores:
Pred. No.: 43.3 Length: 507
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 15 Gaps: 0

US-09-972-032-2 (1-79) x US-10-156-761-6581 (1-507)

Qy 47 AlaSerArgTrpProArgSerAla 54
Db 197 GCATCCCGCTGGCCACGATCCGCG 220

RESULT 6

US-10-282-122A-11225/c
; Sequence 11225, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 11225
; LENGTH: 720
; TYPE: DNA
; ORGANISM: Burkholderia cepacia
US-10-282-122A-11225

Alignment Scores:
Pred. No.: 56.7 Length: 720
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x US-10-282-122A-11225 (1-720)

Qy 47 AlaSerArgTrpProArgSerAla 54
Db 107 GCGAGCCGCTGGCCGCTTCGGCC 84

RESULT 7

US-10-425-114-13766/c
; Sequence 13766, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 13766
; LENGTH: 775
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB143-040-B9_FLI
US-10-425-114-13766

Alignment Scores:
Pred. No.: 59.9 Length: 775
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x US-10-425-114-13766 (1-775)
QY 49 ArgTTPProArgSerAlaSerArg 56
DB 312 CGCTGGCCGCGGTCTGGCGTCGAGG 289
RESULT 8
US-10-369-493-35593/c
; Sequence 35593, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 35593
; LENGTH: 777
; TYPE: DNA
; ORGANISM: Mesorhizobium loti
US-10-369-493-35593
Alignment Scores: 777
Pred. No.: 60.1 Length: 777
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 16 Gaps: 0
US-09-972-032-2 (1-79) x US-10-369-493-35593 (1-777)
QY 49 ArgTTPProArgSerAlaSerArg 56
DB 649 CGATGGCCTCGATCTGCGTCGCGC 626
RESULT 9
US-10-369-493-37767
; Sequence 37767, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 37767
; LENGTH: 969
; TYPE: DNA
; ORGANISM: Pseudomonas fluorescens
US-10-369-493-37767
Alignment Scores: 969
Pred. No.: 71.1 Length: 969
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 10.13% Indels: 0
DB: 16 Gaps: 0
US-09-972-032-2 (1-79) x US-10-369-493-37767 (1-969)
QY 47 AlaSerArgTTPProArgSerAla 54
DB 617 GCAAGCCGCTGGCCAGGAGCGCT 640
RESULT 10
US-10-437-963-38330
; Sequence 38330, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 38330
; LENGTH: 975
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_41977C.1
US-10-437-963-38330
Alignment Scores: 975
Pred. No.: 71.5 Length: 975
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 17 Gaps: 0
US-09-972-032-2 (1-79) x US-10-437-963-38330 (1-975)
QY 49 ArgTTPProArgSerAlaSerArg 56
DB 329 AGGTGGCCTCGGTGGCTAGCCGA 352
RESULT 11
US-10-425-114-21611
; Sequence 21611, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 21611
; LENGTH: 1007
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3354-046-G9_FLI

US-10-425-114-21611

Alignment Scores:
Pred. No.: 73.3 Length: 1007
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x US-10-425-114-21611 (1-1007)

Qy 53 SerAlaSerArgTrpProTrpSer 60
Db 309 TCAGCCTCGCGCTGGCCATGGAGC 332

RESULT 12

US-10-282-122A-13444/c
; Sequence 13444, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13444
; LENGTH: 1035
; TYPE: DNA
; ORGANISM: Burkholderia fungorum

Alignment Scores:
Pred. No.: 74.8 Length: 1035
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x US-10-282-122A-13444 (1-1035)

Qy 48 SerArgTrpProArgSerAlaSer 55
Db 836 TCGAGATGCCGAGGTCGTGCCAGT 813

RESULT 13

US-09-773-926-8/c
; Sequence 8, Application US/09773926
; Patent No. US20020042057A1
; GENERAL INFORMATION:
; APPLICANT: KNOLL, RALPH
; TITLE OF INVENTION: MLP-GENE, NUCLEIC ACIDS, POLYPEPTIDES AND USE THEREOF
; FILE REFERENCE: SCH-1746
; CURRENT APPLICATION NUMBER: US/09/773,926
; CURRENT FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/181,928
; PRIOR FILING DATE: 2000-02-11
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 1037
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-773-926-8

Alignment Scores:

Pred. No.: 74.9 Length: 1037
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 9 Gaps: 0

US-09-972-032-2 (1-79) x US-09-773-926-8 (1-1037)

Qy 26 LeuGlyGluGlyGlnGluGlyGly 33
Db 910 CTGGGGGAGGCGAGGAGGAGGC 887

RESULT 14

US-10-021-723A-7
; Sequence 7, Application US/10021723A
; Publication No. US20030101476A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Mathur, Eric
; APPLICANT: Richardson, Toby
; APPLICANT: Robertson, Dan
; APPLICANT: Barton, Nelson
; TITLE OF INVENTION: Recombinant Phytases and Uses Thereof
; FILE REFERENCE: 112766.140 (DIV-016CIP)
; CURRENT APPLICATION NUMBER: US/10/021,723A
; CURRENT FILING DATE: 2002-10-23
; PRIOR APPLICATION NUMBER: US 60/255,090
; PRIOR FILING DATE: 2000-12-12
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1230
; TYPE: DNA
; ORGANISM: Rhizobium

US-10-021-723A-7

Alignment Scores:

Pred. No.: 85.4 Length: 1230
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 15 Gaps: 0

US-09-972-032-2 (1-79) x US-10-021-723A-7 (1-1230)

Qy 47 AlaSerArgTrpProArgSerAla 54
Db 725 GCAAGCCGATGCCGCGAGTCGGCT 748

RESULT 15

US-10-282-122A-13867
; Sequence 13867, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13867
; LENGTH: 1239
; TYPE: DNA
; ORGANISM: Burkholderia mallei
US-10-282-122A-13867

Alignment Scores:
Pred. No.: 85.9 Length: 1239
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 13 Gaps: 0

US-09-972-032-2 (1-79) x US-10-282-122A-13867 (1-1239)

Qy 48 SerArgTrpProArgSerAlaSer 55
Db 95 AGCAGATGCCGCGAGCGCTTCA 118

Search completed: July 12, 2004, 21:27:05
Job time : 2948 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 12, 2004, 18:17:11 ; Search time 72 Seconds

(without alignments)
608,904 Million cell updates/sec

Title: US-09-972-032-2

Perfect score: 79

Sequence: 1 MCGRPVRSAGCGFADAHWT.....SAGLTVRDRPQLGELCMGRG 79

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 582709 seqs, 277475446 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1359572

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09972032/runat_06072004_121535_8903/app_query.fasta_1.263
-DB=Issued Patents NA -QPMF=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORF=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09972032@cgn 1.1.69 @runat_06072004_121535_8903 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	9	11.4	1402	1	US-08-447-965A-1
2	8	10.1	1335	2	US-08-107-876-25
3	8	10.1	1335	4	US-09-295-820-25
C 4	8	10.1	1539	4	US-09-489-039A-5665
5	8	10.1	1608	4	US-09-252-991A-5081
6	8	10.1	1617	4	US-09-252-991A-5051
C 7	8	10.1	1914	4	US-09-291-922-19
8	8	10.1	2283	4	US-09-252-991A-15346
C 9	8	10.1	2463	4	US-09-252-991A-15430
10	8	10.1	2784	4	US-09-252-991A-15252
C 11	8	10.1	4403765	3	US-09-103-840A-2
C 12	8	10.1	4411529	3	US-09-103-840A-1

13	7	8.9	53	3	US-09-344-888A-3
14	7	8.9	223	6	524792-9
C 15	7	8.9	262	4	US-09-280-116-53
16	7	8.9	291	4	US-09-313-294A-4498
C 17	7	8.9	295	4	US-09-313-294A-4599
18	7	8.9	315	4	US-09-382-552-36
C 19	7	8.9	378	4	US-09-252-991A-15117
20	7	8.9	392	4	US-09-621-976-14323
21	7	8.9	401	4	US-09-621-976-12312
C 22	7	8.9	432	4	US-09-252-991A-10246
23	7	8.9	437	4	US-09-621-976-1458
C 24	7	8.9	471	4	US-09-252-991A-14813
25	7	8.9	471	4	US-09-621-976-1982
26	7	8.9	476	4	US-09-621-976-15482
27	7	8.9	477	4	US-09-252-991A-6345
28	7	8.9	501	4	US-09-252-991A-4653
29	7	8.9	510	4	US-09-252-991A-2991
C 30	7	8.9	543	4	US-09-252-991A-16094
C 31	7	8.9	546	4	US-09-621-976-2743
32	7	8.9	558	4	US-09-621-976-697
33	7	8.9	561	4	US-09-252-991A-368
34	7	8.9	567	4	US-09-252-991A-14500
C 35	7	8.9	582	4	US-09-252-991A-14858
C 36	7	8.9	663	4	US-09-280-116-465
37	7	8.9	678	4	US-09-252-991A-168
38	7	8.9	678	4	US-09-252-991A-523
39	7	8.9	699	4	US-09-252-991A-7908
C 40	7	8.9	765	4	US-09-489-039A-6578
C 41	7	8.9	771	4	US-09-252-991A-597
C 42	7	8.9	789	4	US-09-252-991A-9465
C 43	7	8.9	825	4	US-09-252-991A-5418
C 44	7	8.9	834	4	US-09-252-991A-7643
45	7	8.9	840	4	US-09-252-991A-1634

ALIGNMENTS

RESULT 1

US-08-447-965A-1/c
; Sequence 1, Application US/08447965A
; Patent No. 5776692
; GENERAL INFORMATION:
; APPLICANT: El-Zaatari, Fouad A.K.
; APPLICANT: Graham, David Y.
; APPLICANT: Naser, Saleh
; TITLE OF INVENTION: Mycobacterial Genus-Specific DNA Probe
; TITLE OF INVENTION: and its Expressed Product
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESS: Fulbright & Jaworski L.L.P.
; STREET: 801 Pennsylvania Ave. N.W.
; CITY: Washington, D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,965A
; FILING DATE: 23-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Davis, Peter J.
; REGISTRATION NUMBER: 36,119
; REFERENCE/DOCKET NUMBER: 311.019
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-662-4669
; TELEFAX: 202-662-4643
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1402 base pairs

Sequence 3, Appli
Patent No. 524792
Sequence 53, Appl
Sequence 4498, Ap
Sequence 5899, Ap
Sequence 36, Appl
Sequence 15117, A
Sequence 14323, A
Sequence 12312, A
Sequence 10246, A
Sequence 1458, Ap
Sequence 14813, A
Sequence 1982, Ap
Sequence 15482, A
Sequence 6345, Ap
Sequence 6453, Ap
Sequence 2891, Ap
Sequence 16094, A
Sequence 2743, Ap
Sequence 697, App
Sequence 368, App
Sequence 14500, A
Sequence 14858, A
Sequence 265, App
Sequence 168, App
Sequence 523, App
Sequence 7908, Ap
Sequence 6578, Ap
Sequence 597, App
Sequence 9465, Ap
Sequence 5418, Ap
Sequence 7643, Ap
Sequence 1634, Ap

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
US-08-447-965A-1

Alignment Scores:
Pred. No.: 10.2 Length: 1402
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 11.39% Indels: 0
DB: 1 Gaps: 0

US-09-972-032-2 (1-79) x US-08-447-965A-1 (1-1402)

Qy 47 AlaserArgTrpProArgSerAlaser 55
Db 1260 GCGTCCGTTGGCCAGATCGGCAGT 1234

RESULT 2

US-08-107-676-25
; Sequence 25, Application US/08107676
; Patent No. 5955356
; GENERAL INFORMATION:
; APPLICANT: Content, Jean
; APPLICANT: DeWit, Lucas
; APPLICANT: DeBruyn, Jacqueline
; TITLE OF INVENTION: Mycobacterium Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding Them for Diagnosis and Control of
; TITLE OF INVENTION: Tuberculosis
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 5955356west Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/107,676
; FILING DATE: 04-OCT-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/EP92/00268
; FILING DATE: 07-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91400388.4
; FILING DATE: 14-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalchuk, Katherine M.
; REGISTRATION NUMBER: 36,848
; REFERENCE/DOCKET NUMBER: 8076.89USWO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-9081
; TELEFAX: 612-332-9081
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1335 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium kansasii
; IMMEDIATE SOURCE:
; CLONE: Antigen 85B from M. kansasii

US-08-107-676-25

Alignment Scores:
Pred. No.: 82.4 Length: 1335
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 2 Gaps: 0

US-09-972-032-2 (1-79) x US-08-107-676-25 (1-1335)

Qy 51 ProArgSerAlaserArgTrpPro 58
Db 751 CCGGTCGCGCATCTCGATGCCG 774

RESULT 3

US-09-295-820-25
; Sequence 25, Application US/09295820
; Patent No. 6638511
; GENERAL INFORMATION:
; APPLICANT: Content, Jean
; APPLICANT: DeWit, Lucas
; APPLICANT: DeBruyn, Jacqueline
; TITLE OF INVENTION: Mycobacterium Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding Them for Diagnosis and Control of
; TITLE OF INVENTION: Tuberculosis
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant & Gould
; STREET: 3100 No. 6638511west Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/295,820
; FILING DATE: 21-Apr-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/107,676
; FILING DATE: <Unknown>
; APPLICATION NUMBER: EP 91400388.4
; FILING DATE: 14-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalchuk, Katherine M.
; REGISTRATION NUMBER: 36,848
; REFERENCE/DOCKET NUMBER: 8076.89USWO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1335 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Mycobacterium kansasii
; IMMEDIATE SOURCE:
; CLONE: Antigen 85B from M. kansasii
; SEQUENCE DESCRIPTION: SEQ ID NO: 25:
US-09-295-820-25

Alignment Scores:
Pred. No.: 82.4 Length: 1335
Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 4 Gaps: 0

US-09-972-032-2 (1-79) x US-09-295-820-25 (1-1335)

QY 51 ProArgSerAlaSerArgTrpPro 58
DB 751 CGCGGTGCGCATCTCGATGGCG 774

RESULT 4

US-09-489-039A-5665/c
; Sequence 5665, Application US/09489039A
; Patent No. 6610836

GENERAL INFORMATION:

; APPLICANT: Gary Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; PRIOR FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 5665

; LENGTH: 1539

; TYPE: DNA

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-5665

Alignment Scores:
Pred. No.: 93.5 Length: 1539
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 4 Gaps: 0

US-09-972-032-2 (1-79) x US-09-489-039A-5665 (1-1539)

QY 48 SerArgTrpProArgSerAlaSer 55
DB 1162 AGCAGATGGCGCGGTGAGTTCA 1139

RESULT 5

US-09-252-991A-5081
; Sequence 5081, Application US/09252991A
; Patent No. 6551795

GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 5081

; LENGTH: 1608

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-5081

Alignment Scores:
Pred. No.: 97.2 Length: 1608
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0

DB: 4 Gaps: 0
US-09-972-032-2 (1-79) x US-09-252-991A-5081 (1-1608)

QY 48 SerArgTrpProArgSerAlaSer 55
DB 668 TCGAGGTGGCGCGGTGAGTTCA 691

RESULT 6

US-09-252-991A-5051
; Sequence 5051, Application US/09252991A
; Patent No. 6551795

GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 5051

; LENGTH: 1617

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-5051

Alignment Scores:
Pred. No.: 97.6 Length: 1617
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 4 Gaps: 0

US-09-972-032-2 (1-79) x US-09-252-991A-5051 (1-1617)

QY 48 SerArgTrpProArgSerAlaSer 55
DB 393 TCGAGGTGGCGCGGTGAGTTCA 416

RESULT 7

US-09-291-922-19/c
; Sequence 19, Application US/09291922
; Patent No. 6383776

GENERAL INFORMATION:

; APPLICANT: Allen, Steve

; APPLICANT: Hitz, Bill

; APPLICANT: Kinney, Tony

; APPLICANT: Tingey, Scott

; TITLE OF INVENTION: Plant Sugar Transport Proteins

; FILE REFERENCE: BB-1163

; CURRENT APPLICATION NUMBER: US/09/291,922

; CURRENT FILING DATE: 1999-04-14

; EARLIER APPLICATION NUMBER: 60/083,044

; EARLIER FILING DATE: April 24, 1998

; NUMBER OF SEQ ID NOS: 30

; SOFTWARE: Microsoft Office 97

; SEQ ID NO 19

; LENGTH: 1914

; TYPE: DNA

; ORGANISM: Zea mays

US-09-291-922-19

Alignment Scores:
Pred. No.: 113 Length: 1914
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0

```
DB:                               4          Gaps: 0
US-09-972-032-2 (1-79) x US-09-291-922-19 (1-1914)
QY      25 GlyLeuGlyCluGlyGlnGly 32
DB      1393 GGCCTTGGAGGGGACAGGAAGT 1370

RESULT 8
US-09-252-991A-15346
; Sequence 15346, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15346
; LENGTH: 2283
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15346
Alignment Scores:
Pred. No.: 133          Length: 2283
Score: 8.00            Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 4          Gaps: 0

US-09-972-032-2 (1-79) x US-09-252-991A-15346 (1-2283)
QY      54 AlaSerArgTrpProTrpSerAla 61
DB      1218 GCTTCGAGGTGGCCCTGGAGCGCA 1241

RESULT 9
US-09-252-991A-15430/c
; Sequence 15430, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15430
; LENGTH: 2463
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15430
Alignment Scores:
Pred. No.: 142          Length: 2463
Score: 8.00            Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 4          Gaps: 0

DB:                               4          Gaps: 0
US-09-972-032-2 (1-79) x US-09-252-991A-15430 (1-2463)
QY      54 AlaSerArgTrpProTrpSerAla 61
DB      1073 GCTTCGAGGTGGCCCTGGAGCGCA 1050

RESULT 10
US-09-252-991A-15252
; Sequence 15252, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 15252
; LENGTH: 2784
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-15252
Alignment Scores:
Pred. No.: 158          Length: 2784
Score: 8.00            Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.13% Indels: 0
DB: 4          Gaps: 0

US-09-972-032-2 (1-79) x US-09-252-991A-15252 (1-2784)
QY      54 AlaSerArgTrpProTrpSerAla 61
DB      1787 GCTTCGAGGTGGCCCTGGAGCGCA 1810

RESULT 11
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2
Alignment Scores:
Pred. No.: 1.05e+05      Length: 4403765
Score: 8.00            Matches: 8
Percent Similarity: 100.00% Conservative: 0
```


Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.13% Indels: 0
 DB: Gaps: 0

US-09-972-032-2 (1-79) x US-09-103-840A-2 (1-4403765)

QY 51 ProArgSerAlaSerArgTrpPro 58
 Db 943238 CCGCGTCCGATCGCGCTGGCGG 943215

RESULT 12

US-09-103-840A-1/c
 ; Sequence 1, Application US/09103840A

; Patent No. 6294328
 ; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-20007.00

; CURRENT APPLICATION NUMBER: US/09/103,840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 4411529

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; OTHER INFORMATION: H37Rv

US-09-103-840A-1

Alignment Scores:

Pred. No.: 1.05e+05 Length: 4411529
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.13% Indels: 0
 DB: Gaps: 0

US-09-972-032-2 (1-79) x US-09-103-840A-1 (1-4411529)

QY 51 ProArgSerAlaSerArgTrpPro 58
 Db 943348 CCGCGTCCGATCGCGCTGGCGG 943325

RESULT 13

US-09-344-888A-3

; Sequence 3, Application US/09344888A

; Patent No. 6291245

; GENERAL INFORMATION:

; APPLICANT: Kopetzki, Erhard

; APPLICANT: Schantz, Christian

; TITLE OF INVENTION: New Host-Vector System

; FILE REFERENCE: CD20315

; CURRENT APPLICATION NUMBER: US/09/344,888A

; CURRENT FILING DATE: 1999-06-25

; PRIOR FILING DATE: 1998-07-15

; PRIOR FILING DATE: 1998-07-15

; PRIOR FILING DATE: 1998-10-09

; NUMBER OF SEQ ID NOS: 24

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3

; LENGTH: 53

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:primer

US-09-344-888A-3

Alignment Scores:

Alignment Scores:
 Pred. No.: 165 Length: 262
 Score: 7.00 Matches: 7

Pred. No.: 39.9 Length: 53
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.86% Indels: 0
 DB: Gaps: 0

US-09-972-032-2 (1-79) x US-09-344-888A-3 (1-53)

QY 27 GlyGlyGlnGluGlyGly 33
 Db 32 GGAGAGGGCCAGAGGAGGG 52

RESULT 14

5244792-9

; Patent No. 5244792

; APPLICANT: BURKE, RAE L.; PACHL, CAROL; VALENZUELA, PABLO D.T.

; TITLE OF INVENTION: EXPRESSION OF RECOMBINANT GLYOPIROTEIN

; B FROM HERPES SIMPLEX VIRUS

; NUMBER OF SEQUENCES: 19

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/587,179

; FILING DATE: 20-SEP-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 921,730

; FILING DATE: 20-OCT-1986

; APPLICATION NUMBER: 597,784

; FILING DATE: 06-APR-1984

; SEQ ID NO:9

; LENGTH: 223

5244792-9

Alignment Scores:

Pred. No.: 143 Length: 223
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.86% Indels: 0
 DB: Gaps: 0

US-09-972-032-2 (1-79) x 5244792-9 (1-223)

QY 50 TrpProArgSerAlaSerArg 56
 Db 202 TGGCGCGTGGCGTGGCG 222

RESULT 15

US-09-280-116-53/c

; Sequence 53, Application US/09280116A

; Patent No. 6331427

; GENERAL INFORMATION:

; APPLICANT: Robison, Keith E.

; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Human Protease Homologs

; FILE REFERENCE: 5800-24, 035800/176965

; CURRENT APPLICATION NUMBER: US/09/280,116A

; CURRENT FILING DATE: 1999-03-26

; NUMBER OF SEQ ID NOS: 268

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 53

; LENGTH: 262

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: aspartyl proteases

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)-(262)

; OTHER INFORMATION: n = a, t, c or g

US-09-280-116-53

Alignment Scores:

Alignment Scores:
 Pred. No.: 165 Length: 262
 Score: 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 8.86% Indels: 0
 DB: 4 Gaps: 0

US-09-972-032-2 (1-79) x US-09-280-116-53 (1-262)

QY 69 ProGlnLeuGlyGluLeuCys 75
 ||||||||||||||||
 Db 169 CCCCACTTGGGGAGCTCTGC 149

Search completed: July 12, 2004, 18:46:21
 Job time : 1734 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2004, 11:56:49 ; Search time 495 Seconds
(without alignments)
8496.396 Million cell updates/sec

Title: US-09-972-032-1
Perfect score: 990
Sequence: 1 ggaatgtctcaggagccaa.....aaaatgagcgccgaagt 990

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: Geneseqn1990s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	990	100.0	990	6	ABL60606 Human ERC
2	147	14.8	597	5	AAS69040 DNA encod
3	80	8.1	404	4	AAL01097 Human rep
4	80	8.1	404	4	ABL96563 Human tes
5	51	5.2	149480	6	ABL61947 Colon ade
6	51	5.2	149480	6	ABL68365 Kidney ca
7	51	5.2	149480	6	ABL61948 Colon ade
8	48	4.8	118584	9	ADC87623 Human GPC
9	48	4.8	349999	9	ADC87010 Human GPC
10	47	4.7	494	5	ABA15806 Human ner
11	47	4.7	9888	4	AAK71311 Human imm
12	47	4.7	13160	5	ABA15804 Human ner
13	47	4.7	13189	5	ABA15805 Human ner
14	47	4.7	43419	9	ADC86998 Human GPC
15	47	4.7	63626	9	ADC27000 Human car
16	47	4.7	63720	8	ADA03080 Human PRD
17	47	4.7	63720	8	ADA66364 Human PRD
18	47	4.7	63720	8	ADB72818 Human PRD
19	47	4.7	103747	6	ABQ86139 Human ost
20	47	4.7	122748	6	ABT10719 Human bre
21	46	4.6	274	3	AAC19185 Human sec
22	46	4.6	394	4	AAI82243 Human pol
23	46	4.6	2105	4	AAH99495 Human pro

24	46	4.6	2904	9	ADB63202	Adb63202 Human CDN
25	46	4.6	2940	8	AAD58740	Aad58740 Human tra
c	26	4.6	8865	4	AAK87029	Aak87029 Human imm
27	46	4.6	12117	3	AA96368	AA96368 Polymorph
28	46	4.6	22452	4	AAS27829	Aas27829 DNA encod
29	46	4.6	22452	4	AAS27827	Aas27827 DNA encod
30	46	4.6	22452	9	ADB94632	ADB94632 Novel hum
31	46	4.6	22452	9	ADB94630	ADB94630 Novel hum
32	46	4.6	32193	4	AAI16595	AAI16595 Human nov
33	46	4.6	32193	4	AAI36258	AAI36258 Human mus
34	46	4.6	32193	7	ABX59246	ABX59246 cDNA enco
35	46	4.6	32249	4	AAL04676	Aal04676 Human rep
36	46	4.6	32249	4	ABL97583	Abi97583 Human tes
c	37	46	50000	3	AA96363	AA96363 Polymorph
38	46	4.6	75270	7	ACA63030	ACA63030 Human chr
39	46	4.6	75270	9	ADD71050	Add71050 Human sec
40	45	4.5	122	4	AAK84551	Aak84551 Human imm
c	41	45	308	4	AA332644	Aas332644 Human gen
c	42	45	354	4	AA337387	Aas337387 Novel hum
c	43	45	414	4	AAI89811	Aai89811 Human pol
c	44	45	444	5	ABV53291	Abv53291 Human pro
45	45	4.5	482	4	AAI91309	Aai91309 Human pol

ALIGNMENTS

RESULT 1

ABL60606
ID ABL60606 standard; cDNA; 990 BP.

XX ABL60606;

DT 27-AUG-2002 (first entry)

XX Human ERCoA3 protein encoding cDNA.

KW Estrogen Receptor Coregulator 3; ERCoA3; tamoxifen; estrogen; cancer;

KW osteoporosis; cytostatic; osteopathic; human; gene; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 203..442

FT /*tag= a

FT /product= "ERCoA3 protein"

XX WO200228352-A2.

XX 11-APR-2002.

XX 05-OCT-2001; 2001WO-US031271.

XX 05-OCT-2000; 2000US-0238190P.

XX (UYCA-) UNIV CASE WESTERN RESERVE.

XX Montano M, Sutton A;

XX WPI; 2002-454492/48.

XX P-PSDB; ABB08035.

XX New polypeptide, that is a functional equivalent of ERCoA3 (Estrogen Receptor Coregulator 3), is useful in inhibiting or reducing tamoxifen or estrogen-induced proliferation of cancer cells and in treating osteoporosis.

XX Claim 5; Fig 1; 39pp; English.

XX The invention relates to a ERCoA3 (Estrogen Receptor Coregulator 3) protein and encoding polynucleotides. ERCoA3 can be used to inhibit or reduce tamoxifen or estrogen induced proliferation of cancer cells, by reduced activity of ERCoA3, and for detecting cancer cells that are

CC tamoxifen resistant, or to treat osteoporosis, by increasing levels of
CC ERCoA3 in cells. The encoding polynucleotide can be used to inhibit
CC translation of a mRNA encoding ERCoA3. ERCoA3 acts as a coregulator
CC protein and can bind to the estrogen receptor to activate a molecular
CC cellular response in the cell. The present sequence represents the human
CC ERCoA3 encoding cDNA
XX
SQ Sequence 990 BP; 189 A; 281 C; 317 G; 203 T; 0 U; 0 Other;

Query Match 100.0%; Score 990; DB 6; Length 990;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 990; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGAATTGTTCTCGAGGCCAAGAAATTCGGCAAGCGCATCTCTGACCTCGTGATCCGCCCG 60
DB 1 GGAATTGTTCTCGAGGCCAAGAAATTCGGCAAGCGCATCTCTGACCTCGTGATCCGCCCG 60

QY 61 CCTCAGCTTCCCAAGTCTGGGATTCAGGCGATGAGCCACCGCGCCCGCCCGCCGACAC 120
DB 61 CCTCAGCTTCCCAAGTCTGGGATTCAGGCGATGAGCCACCGCGCCCGCCCGCCGACAC 120

QY 121 CTAGTTTAAAGGCGCCCTCTGTTGCTGCGCTGCGCGCTCCCGAGTGCCTGCTG 180
DB 121 CTAGTTTAAAGGCGCCCTCTGTTGCTGCGCTGCGCGCTCCCGAGTGCCTGCTG 180

QY 181 GCGGGCTCAGTCCCGGTTGCCATGTTGGAGACCGCGTTCGCTGAAGCGCTGGATGG 240
DB 181 GCGGGCTCAGTCCCGGTTGCCATGTTGGAGACCGCGTTCGCTGAAGCGCTGGATGG 240

QY 241 CTTGCGTGATGCATGACATGGACCGGCTCTGAGTGGCTAGGAGAGGACGAGGGCGG 300
DB 241 CTTGCGTGATGCATGACATGGACCGGCTCTGAGTGGCTAGGAGAGGACGAGGGCGG 300

QY 301 AATTGGCCCCAGGGCCAGTCTGCGCCAGACCCCGACTGCGCTCCCGTGCCTCCGCGAG 360
DB 301 AATTGGCCCCAGGGCCAGTCTGCGCCAGACCCCGACTGCGCTCCCGTGCCTCCGCGAG 360

QY 361 CGCTCCCGTGGCTGGAGTGCAGTCTTACCGTCCGAGATCTGCTCCCACTGGCGA 420
DB 361 CGCTCCCGTGGCTGGAGTGCAGTCTTACCGTCCGAGATCTGCTCCCACTGGCGA 420

QY 421 GCTGTGATCGGCGTGGCTTAAGCGCTGTTGGTTAGATTGGCCAGCGGACTTAAG 480
DB 421 GCTGTGATCGGCGTGGCTTAAGCGCTGTTGGTTAGATTGGCCAGCGGACTTAAG 480

QY 481 TGTGTCTGTAAGAGCATGGAATTAAGTGTGAGGGTCTGGAAGAGTATCCCGCC 540
DB 481 TGTGTCTGTAAGAGCATGGAATTAAGTGTGAGGGTCTGGAAGAGTATCCCGCC 540

QY 541 CACATCAATGCGCTTAGTCTAGGAAGCGGCTGGTGGGGCTTTAGGGCGAGGGG 600
DB 541 CACATCAATGCGCTTAGTCTAGGAAGCGGCTGGTGGGGCTTTAGGGCGAGGGG 600

QY 601 CAGACATACCCCAAGTGTGATTTATACCGCAAGGGCTGGATCGAACCCCGCCAA 660
DB 601 CAGACATACCCCAAGTGTGATTTATACCGCAAGGGCTGGATCGAACCCCGCCAA 660

QY 661 GACATGGAGGCTGTGCTGCTGAGAGGCGCGGCAATCCAGTGTGCTGGCTTTAC 720
DB 661 GACATGGAGGCTGTGCTGCTGAGAGGCGCGGCAATCCAGTGTGCTGGCTTTAC 720

QY 721 AGAAAGAGCTCCACCTTCTGAGTGTGAGATGCGATCTAGTGTGCTCCACCGATGG 780
DB 721 AGAAAGAGCTCCACCTTCTGAGTGTGAGATGCGATCTAGTGTGCTCCACCGATGG 780

QY 781 GAGCTGGGGCGGGGAGATGCTGCCCGAGTACAAAGCTGATTGACCTGGGGCTCTG 840
DB 781 GAGCTGGGGCGGGGAGATGCTGCCCGAGTACAAAGCTGATTGACCTGGGGCTCTG 840

QY 841 GACTTCCCTGATCTCTGCTTGCATCTCCAGCAAGTCTCTGCTCCCGTGGCTGCCTTAT 900
DB 841 GACTTCCCTGATCTCTGCTTGCATCTCCAGCAAGTCTCTGCTCCCGTGGCTGCCTTAT 900

RESULT 2
AAS69040/c
ID AAS69040 standard; cDNA; 597 BP.
XX
AC AAS69040;
DT 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #4844.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.
KW Homo sapiens.
OS WO200175067-A2.
PN 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US008631.
PF 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR F-PSDB; ABG04853.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX Claim 1; SEQ ID NO 4844; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities. The
CC polypeptide and polynucleotide sequences have applications in:
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 597 BP; 97 A; 217 C; 196 G; 87 T; 0 U; 0 Other;
Query Match 14.8%; Score 147; DB 5; Length 597;
Best Local Similarity 100.0%; Pred. No. 4e-48;

PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251388P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-02559678P.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
DR P-PSDB; AAM95127.
XX PT Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.
XX PS Claim 1; SEQ ID NO 1098; 1297pp + Sequence Listing; English.
XX CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a coding sequence of the
CC invention
XX SQ Sequence 404 BP; 83 A; 109 C; 107 G; 95 T; 0 U; 10 Other;
Query Match 8.1%; Score 80; DB 4; Length 404;
Best Local Similarity 100.0%; Pred. No. 8.3e-22;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 872 CAAAGTCCTGCGGTGGTGGCTTCACACCTCTCTCACTTCTGCTTCAGAGTAA 931
Db 8 CAAAGTCCTGCGGTGGTGGCTTCACACCTCTCTCACTTCTGCTTCAGAGTAA 67
QY 932 AATTGCAAGATCTGTGTGC 951
Db 68 AATTGCAAGATCTGTGTGC 87
RESULT 4
ABL96563
ID ABL96563 standard; cDNA; 404 BP.
XX AC ABL96563;
XX DT 21-JUN-2002 (first entry)
XX DE Human testicular antigen encoding cDNA SEQ ID NO: 231.
XX KW Human; testicular antigen; testes; cancer; metastasis; immune disorder;
KW reproductive system disorder; urinary system disorder; gene therapy;
KW cardiovascular disorder; respiratory disorder; neurological disorder;

KW gastrointestinal disease; infection; cytostatic; gene; ss.
XX Homo sapiens.
CS WO200155317-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US001329.
XX PR 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184664P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225477P.
PR 14-AUG-2000; 2000US-0225577P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.

```
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 28-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 29-SEP-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239335P.
PR 13-OCT-2000; 2000US-0239393P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250331P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251858P.
PR 08-DEC-2000; 2000US-0251859P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-483232/52.
XX Nucleic acids encoding 973 human testicular antigen polypeptides, useful
PT for preventing, diagnosing and/or treating testicular cancer.
XX Claim 1; SEQ ID NO 231; 766pp; English.
XX The present invention provides the protein and coding sequences of 973
CC human testicular antigens, and fragments of their genomic sequences. The
CC sequences can be used in the treatment of cardiovascular, urinary system,
CC reproductive system, immune, respiratory, neurological and
CC gastrointestinal disorders, infections, and particularly cancer,
CC especially testicular cancers. The present sequence is a cDNA of the
CC invention
XX
SQ Sequence 404 BP; 83 A; 109 C; 107 G; 95 T; 0 U; 10 Other;

Query Match 8.1%; Score 80; DB 4; Length 404;
Best Local Similarity 100.0%; Pred. No. 8.3e-22;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 872 CAAGTCCTGTCCCGTGGCTTCATCCACTCTCTCACTTCTGCTTCCAGTAA 931
Db |||||||
QY 932 AATTGCAAGATCTGTGGTGC 951
Db |||||||
68 AATTGCAAGATCTGTGGTGC 87

RESULT 5
ABL61947
ID ABL61947 standard; DNA; 149480 BP.
XX
AC ABL61947;
XX
DT 15-MAY-2002 (first entry)
XX
DE Colon adenocarcinoma related gene sequence SEQ ID NO:284.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW Gene; ds.
XX
OS Homo sapiens.
XX
PN WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US010838.
XX
PR 05-JUN-2000; 2000US-0209473P.
PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-0233133P.
PR 18-SEP-2000; 2000US-0233617P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235280P.
```

PR 26-SEP-2000; 2000US-0235637P.
 PR 26-SEP-2000; 2000US-0235638P.
 PR 27-SEP-2000; 2000US-0235711P.
 PR 27-SEP-2000; 2000US-0235720P.
 PR 27-SEP-2000; 2000US-0235840P.
 PR 27-SEP-2000; 2000US-0235863P.
 PR 28-SEP-2000; 2000US-0236028P.
 PR 28-SEP-2000; 2000US-0236032P.
 PR 28-SEP-2000; 2000US-0236033P.
 PR 28-SEP-2000; 2000US-0236034P.
 PR 28-SEP-2000; 2000US-0236109P.
 PR 28-SEP-2000; 2000US-0236111P.
 PR 29-SEP-2000; 2000US-0236842P.
 PR 29-SEP-2000; 2000US-0236891P.
 PR 02-OCT-2000; 2000US-0237172P.
 PR 02-OCT-2000; 2000US-0237173P.
 PR 02-OCT-2000; 2000US-0237278P.
 PR 02-OCT-2000; 2000US-0237294P.
 PR 02-OCT-2000; 2000US-0237295P.
 PR 03-OCT-2000; 2000US-0237316P.
 PR 03-OCT-2000; 2000US-0237425P.
 PR 03-OCT-2000; 2000US-0237598P.
 PR 03-OCT-2000; 2000US-0237604P.
 PR 03-OCT-2000; 2000US-0237606P.
 PR 03-OCT-2000; 2000US-0237608P.
 PR 01-NOV-2000; 2000US-0244867P.
 PR 01-NOV-2000; 2000US-0245084P.
 XX
 PA (AVALON PHARM.
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 XX WPI; 2002-188264/24.
 XX
 PT Screening for anti-neoplastic agent involves exposing cells to a chemical
 PT agent to be tested for anti-neoplastic activity, and determining a change
 PT in expression of a gene of a signature gene set.
 XX
 PS Claim 1; SEQ ID NO 284; 44pp; English.
 XX
 CC The present invention describes a method (M1) for screening for an anti-
 CC neoplastic agent. The method involves exposing cells to a chemical agent
 CC to be tested for anti-neoplastic activity, determining a change in
 CC expression of at least one gene (I) of a signature gene set, where (I)
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
 CC to ABL70110), or is at least 95% identical to (S), where a change in
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
 CC activity and can be used in gene therapy. M1 can be used for screening an
 CC anti-neoplastic agent, and can be used for producing a product which is
 CC the data collected with respect to the anti-neoplastic agent as a result
 CC of M1, and the data is sufficient to convey the chemical structure and/or
 CC properties of the agent. M1 can be used in the treatment of cancer such
 CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
 CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
 CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
 CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilms
 CC tumour
 XX
 SQ Sequence 149480 BP; 38770 A; 34233 C; 35027 G; 40449 T; 0 U; 1 Other;
 Query Match 5.2%; Score 51; DB 6; Length 149480;
 Best Local Similarity 100.0%; Pred. No. 7.6e-11;
 Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 57 CCGCCTCAGCTTCCCAAGTGTGGATTACAGGATGAGCCACCGGCC 107
 DB 76386 CCGCCTCAGCTTCCCAAGTGTGGATTACAGGATGAGCCACCGGCC 76436
 RESULT 6
 ABL68365
 ID ABL68365 standard; DNA; 149480 BP.

XX ABL68365;
 AC 15-MAY-2002 (first entry)
 DT Kidney cancer related gene sequence SEQ ID NO:6702.
 XX
 DE Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
 KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
 KW cytostatic; gene therapy; antineoplastic; Wilms tumour; adenocarcinoma;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX WO200194629-A2.
 PN 13-DEC-2001.
 XX
 PD 30-MAY-2001; 2001WO-US010838.
 PF 05-JUN-2000; 2000US-0209473P.
 PR 05-JUN-2000; 2000US-0209531P.
 PR 18-SEP-2000; 2000US-0233133P.
 PR 18-SEP-2000; 2000US-0233617P.
 PR 20-SEP-2000; 2000US-0234009P.
 PR 20-SEP-2000; 2000US-0234034P.
 PR 22-SEP-2000; 2000US-0234052P.
 PR 22-SEP-2000; 2000US-0234509P.
 PR 22-SEP-2000; 2000US-0234567P.
 PR 25-SEP-2000; 2000US-0234923P.
 PR 25-SEP-2000; 2000US-0234924P.
 PR 25-SEP-2000; 2000US-0235077P.
 PR 25-SEP-2000; 2000US-0235082P.
 PR 25-SEP-2000; 2000US-0235134P.
 PR 25-SEP-2000; 2000US-0235280P.
 PR 26-SEP-2000; 2000US-0235637P.
 PR 26-SEP-2000; 2000US-0235638P.
 PR 27-SEP-2000; 2000US-0235711P.
 PR 27-SEP-2000; 2000US-0235720P.
 PR 27-SEP-2000; 2000US-0235840P.
 PR 27-SEP-2000; 2000US-0235863P.
 PR 28-SEP-2000; 2000US-0236028P.
 PR 28-SEP-2000; 2000US-0236032P.
 PR 28-SEP-2000; 2000US-0236033P.
 PR 28-SEP-2000; 2000US-0236034P.
 PR 28-SEP-2000; 2000US-0236109P.
 PR 28-SEP-2000; 2000US-0236111P.
 PR 29-SEP-2000; 2000US-0236842P.
 PR 29-SEP-2000; 2000US-0236891P.
 PR 02-OCT-2000; 2000US-0237172P.
 PR 02-OCT-2000; 2000US-0237173P.
 PR 02-OCT-2000; 2000US-0237278P.
 PR 02-OCT-2000; 2000US-0237294P.
 PR 02-OCT-2000; 2000US-0237295P.
 PR 03-OCT-2000; 2000US-0237316P.
 PR 03-OCT-2000; 2000US-0237425P.
 PR 03-OCT-2000; 2000US-0237598P.
 PR 03-OCT-2000; 2000US-0237604P.
 PR 03-OCT-2000; 2000US-0237606P.
 PR 03-OCT-2000; 2000US-0237608P.
 PR 01-NOV-2000; 2000US-0244867P.
 PR 01-NOV-2000; 2000US-0245084P.
 XX
 PA (AVALON PHARM.
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
 PI Soppet DR, Weaver Z;
 XX WPI; 2002-188264/24.
 XX
 PT Screening for anti-neoplastic agent involves exposing cells to a chemical
 PT agent to be tested for anti-neoplastic activity, and determining a change
 PT in expression of a gene of a signature gene set.


```
XX Claim 1; SEQ ID NO 6702; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC tumour
XX
SQ Sequence 149480 BP; 39770 A; 34233 C; 35027 G; 40449 T; 0 U; 1 Other;
Query Match 5.2%; Score 51; DB 6; Length 149480;
Best Local Similarity 100.0%; Pred.No. 7.6e-11;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 57 CCGCGCTCAGCTTCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCC 107
DB 76386 CCGCGCTCAGCTTCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCC 76436
RESULT 7
ABL61948
ID ABL61948 standard; DNA; 149480 BP.
XX
AC ABL61948;
XX
DT 15-MAY-2002 (first entry)
DE Colon adenocarcinoma related gene sequence SEQ ID NO:285.
XX
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW gene; ds.
XX
OS Homo sapiens.
XX
XX WO200194629-A2.
XX
PD 13-DEC-2001.
XX
PF 30-MAY-2001; 2001WO-US010838.
XX
PR 05-JUN-2000; 2000US-0209473P.
PR 05-JUN-2000; 2000US-0209531P.
PR 18-SEP-2000; 2000US-023133P.
PR 18-SEP-2000; 2000US-0233617P.
PR 20-SEP-2000; 2000US-0234009P.
PR 20-SEP-2000; 2000US-0234034P.
PR 20-SEP-2000; 2000US-0234052P.
PR 22-SEP-2000; 2000US-0234509P.
PR 22-SEP-2000; 2000US-0234567P.
PR 25-SEP-2000; 2000US-0234923P.
PR 25-SEP-2000; 2000US-0234924P.
PR 25-SEP-2000; 2000US-0235077P.
PR 25-SEP-2000; 2000US-0235082P.
PR 25-SEP-2000; 2000US-0235134P.
PR 25-SEP-2000; 2000US-0235280P.
PR 26-SEP-2000; 2000US-0235637P.
PR 26-SEP-2000; 2000US-0235638P.
PR 27-SEP-2000; 2000US-0235711P.
```

```
PR 27-SEP-2000; 2000US-0235720P.
PR 27-SEP-2000; 2000US-0235840P.
PR 27-SEP-2000; 2000US-0235863P.
PR 28-SEP-2000; 2000US-0236028P.
PR 28-SEP-2000; 2000US-0236032P.
PR 28-SEP-2000; 2000US-0236033P.
PR 28-SEP-2000; 2000US-0236034P.
PR 28-SEP-2000; 2000US-0236109P.
PR 28-SEP-2000; 2000US-0236111P.
PR 29-SEP-2000; 2000US-0236842P.
PR 29-SEP-2000; 2000US-0236891P.
PR 02-OCT-2000; 2000US-0237172P.
PR 02-OCT-2000; 2000US-0237173P.
PR 02-OCT-2000; 2000US-0237278P.
PR 02-OCT-2000; 2000US-0237294P.
PR 02-OCT-2000; 2000US-0237295P.
PR 03-OCT-2000; 2000US-0237316P.
PR 03-OCT-2000; 2000US-0237425P.
PR 03-OCT-2000; 2000US-0237598P.
PR 03-OCT-2000; 2000US-0237604P.
PR 03-OCT-2000; 2000US-0237606P.
PR 03-OCT-2000; 2000US-0237608P.
PR 01-NOV-2000; 2000US-0244867P.
PR 01-NOV-2000; 2000US-0245084P.
XX
PA (AVAL-) AVALON PHARM.
XX
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI Soppet DR, Weaver Z;
XX WPI; 2002-188264/24.
XX
PT Screening for anti-neoplastic agent involves exposing cells to a chemical
PT agent to be tested for anti-neoplastic activity, and determining a change
PT in expression of a gene of a signature gene set.
XX
XX Claim 1; SEQ ID NO 285; 44pp; English.
XX
CC The present invention describes a method (M1) for screening for an anti-
CC neoplastic agent. The method involves exposing cells to a chemical agent
CC to be tested for anti-neoplastic activity, determining a change in
CC expression of at least one gene (I) of a signature gene set, where (I)
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC to ABL70110), or is at least 95% identical to (S), where a change in
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC activity and can be used in gene therapy. M1 can be used for screening an
CC anti-neoplastic agent, and can be used for producing a product which is
CC the data collected with respect to the anti-neoplastic agent as a result
CC of M1, and the data is sufficient to convey the chemical structure and/or
CC properties of the agent. M1 can be used in the treatment of cancer such
CC as colon, breast, stomach, lung, thyroid, oesophageal, ovarian, kidney,
CC prostate or pancreatic cancer, adenocarcinoma, carcinoma, clear cell
CC cancer, infiltrating ductal cancer, infiltrating lobular cancer, squamous
CC cell carcinoma, neuroendocrine carcinoma, papillary carcinoma and Wilm's
CC tumour
XX
SQ Sequence 149480 BP; 39770 A; 34233 C; 35027 G; 40449 T; 0 U; 1 Other;
Query Match 5.2%; Score 51; DB 6; Length 149480;
Best Local Similarity 100.0%; Pred.No. 7.6e-11;
Matches 51; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 57 CCGCGCTCAGCTTCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCC 107
DB 76386 CCGCGCTCAGCTTCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCC 76436
RESULT 8
ADC87623/c
ID ADC87623 standard; DNA; 118584 BP.
XX
AC ADC87623;
XX
```

DT 01-JAN-2004 (first entry)
 DE Human GPCR related polynucleotide SEQ ID NO:2076.
 XX ds; human; GPCR; guanosine triphosphate-binding protein coupled receptor;
 KW gene therapy.
 XX Homo sapiens.
 OS
 XX EP1270724-A2.
 FN
 XX 02-JAN-2003.
 PD
 XX
 XX 18-JUN-2002; 2002EP-00013517.
 PF
 XX 18-JUN-2001; 2001JP-00246789.
 PR
 XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
 PA Suwa M, Asai K, Akiyama Y, Aburatani H;
 XX WPI; 2003-315783/31.
 DR
 XX New polynucleotide, useful for preparing a composition for treating a
 XX patient in need of increased or suppressed activity or expression of the
 XX guanosine triphosphate-binding protein coupled receptor.
 XX Disclosure; SEQ ID NO 2076; 28pp; English.
 PS
 XX The invention relates to a novel polynucleotide encoding a guanosine
 XX triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
 XX the invention may have a use in gene therapy. The polynucleotide and
 XX polypeptide are useful for preparing a composition for treating a patient
 XX in need of increased or suppressed activity or expression of the
 XX guanosine triphosphate-binding protein coupled receptor. The protein
 XX sequences shown in ADC87618-ADC87623 represent polynucleotide sequences
 XX related to the invention.
 XX
 SQ Sequence 118584 BP; 32219 A; 27585 C; 27306 G; 31474 T; 0 U; 0 Other;
 Query Match 4.8%; Score 48; DB 9; Length 118584;
 Best Local Similarity 100.0%; Pred. No. 1.2e-09;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 PS
 XX 57 CCGCCTCAGCTTCCCAAAGTCTGGATTACAGGCATGAGCCACCGC 104
 QY
 DB 29455 CCGCCTCAGCTTCCCAAAGTCTGGATTACAGGCATGAGCCACCGC 29408
 RESULT 9
 ADC87010/c
 ID ADC87010 standard; DNA; 349999 BP.
 XX
 AC ADC87010;
 XX
 DT 01-JAN-2004 (first entry)
 DE Human GPCR gene SEQ ID NO:1463.
 XX
 XX ds; gene; human; GPCR;
 KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
 XX Homo sapiens.
 OS
 XX EP1270724-A2.
 FN
 XX 02-JAN-2003.
 PD
 XX 18-JUN-2002; 2002EP-00013517.
 PF
 XX 19-JUN-2001; 2001JP-00246789.
 PR
 XX

PA (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 PA (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
 XX Suwa M, Asai K, Akiyama Y, Aburatani H;
 PI WPI; 2003-315783/31.
 DR P-PSDB; ADC87011.
 DR
 XX New polynucleotide, useful for preparing a composition for treating a
 XX patient in need of increased or suppressed activity or expression of the
 XX guanosine triphosphate-binding protein coupled receptor.
 XX Claim 1; SEQ ID NO 1463; 28pp; English.
 PS
 XX The invention relates to a novel polynucleotide encoding a guanosine
 XX triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
 XX the invention may have a use in gene therapy. The polynucleotide and
 XX polypeptide are useful for preparing a composition for treating a patient
 XX in need of increased or suppressed activity or expression of the
 XX guanosine triphosphate-binding protein coupled receptor. The
 XX polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
 XX invention.
 XX
 SQ Sequence 349999 BP; 91105 A; 78863 C; 81174 G; 98857 T; 0 U; 0 Other;
 Query Match 4.8%; Score 48; DB 9; Length 349999;
 Best Local Similarity 100.0%; Pred. No. 9.9e-10;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 PS
 QY 57 CCGCCTCAGCTTCCCAAAGTCTGGATTACAGGCATGAGCCACCGC 104
 DB 329421 CCGCCTCAGCTTCCCAAAGTCTGGATTACAGGCATGAGCCACCGC 329374
 RESULT 10
 ABA15806
 ID ABA15806 standard; DNA; 494 BP.
 XX
 AC ABA15806;
 XX
 DT 23-JAN-2002 (first entry)
 DE Human nervous system related polynucleotide SEQ ID NO 8137.
 XX
 XX Human; nontropic; neuroprotective; cytostatic; dermatological; virucide;
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
 KW antiparkinsonian; antitickling; antianaemic; antiarthritic; cancer;
 KW antineumatic; hepatotropic; cerebroprotective; antinflammatory;
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
 XX Homo sapiens.
 OS
 XX WO200159063-A2.
 PN
 XX 16-AUG-2001.
 PD
 XX
 XX 17-JAN-2001; 2001WO-US001334.
 PF
 XX 31-JAN-2000; 2000US-0179065P.
 PR
 XX 04-FEB-2000; 2000US-0180628P.
 PR
 XX 24-FEB-2000; 2000US-0184664P.
 PR
 XX 02-MAR-2000; 2000US-0186350P.
 PR
 XX 16-MAR-2000; 2000US-0189874P.
 PR
 XX 17-MAR-2000; 2000US-0190076P.
 PR
 XX 18-APR-2000; 2000US-0198123P.
 PR
 XX 19-MAY-2000; 2000US-0205515P.
 PR
 XX 07-JUN-2000; 2000US-0209467P.
 PR
 XX 28-JUN-2000; 2000US-0214886P.
 PR
 XX 30-JUN-2000; 2000US-0215135P.
 PR
 XX 07-JUL-2000; 2000US-0216647P.
 PR
 XX 07-JUL-2000; 2000US-0216880P.

PR	11-JUL-2000;	2000US-0217487P;
PR	11-JUL-2000;	2000US-0217496P;
PR	14-JUL-2000;	2000US-0218290P;
PR	26-JUL-2000;	2000US-0220963P;
PR	26-JUL-2000;	2000US-0220964P;
PR	14-AUG-2000;	2000US-0224518P;
PR	14-AUG-2000;	2000US-0224519P;
PR	14-AUG-2000;	2000US-0225211P;
PR	14-AUG-2000;	2000US-0225214P;
PR	14-AUG-2000;	2000US-0225266P;
PR	14-AUG-2000;	2000US-0225267P;
PR	14-AUG-2000;	2000US-0225270P;
PR	14-AUG-2000;	2000US-0225277P;
PR	14-AUG-2000;	2000US-0225751P;
PR	14-AUG-2000;	2000US-0225758P;
PR	14-AUG-2000;	2000US-0225759P;
PR	18-AUG-2000;	2000US-0226719P;
PR	22-AUG-2000;	2000US-0226681P;
PR	22-AUG-2000;	2000US-0226686P;
PR	22-AUG-2000;	2000US-0227182P;
PR	23-AUG-2000;	2000US-0227009P;
PR	30-AUG-2000;	2000US-0228924P;
PR	01-SEP-2000;	2000US-0228287P;
PR	01-SEP-2000;	2000US-0228343P;
PR	01-SEP-2000;	2000US-0229344P;
PR	01-SEP-2000;	2000US-0229345P;
PR	05-SEP-2000;	2000US-0229530P;
PR	05-SEP-2000;	2000US-0229513P;
PR	06-SEP-2000;	2000US-0230436P;
PR	06-SEP-2000;	2000US-0230437P;
PR	08-SEP-2000;	2000US-0231242P;
PR	08-SEP-2000;	2000US-0231243P;
PR	08-SEP-2000;	2000US-0231244P;
PR	08-SEP-2000;	2000US-0231413P;
PR	08-SEP-2000;	2000US-0231414P;
PR	08-SEP-2000;	2000US-0232080P;
PR	08-SEP-2000;	2000US-0232081P;
PR	12-SEP-2000;	2000US-0231968P;
PR	14-SEP-2000;	2000US-0233297P;
PR	14-SEP-2000;	2000US-0233298P;
PR	14-SEP-2000;	2000US-0233399P;
PR	14-SEP-2000;	2000US-0233400P;
PR	14-SEP-2000;	2000US-0233401P;
PR	14-SEP-2000;	2000US-0233403P;
PR	14-SEP-2000;	2000US-0233406P;
PR	14-SEP-2000;	2000US-0233405P;
PR	21-SEP-2000;	2000US-0234223P;
PR	21-SEP-2000;	2000US-0234274P;
PR	25-SEP-2000;	2000US-0234997P;
PR	25-SEP-2000;	2000US-0234998P;
PR	26-SEP-2000;	2000US-0234584P;
PR	29-SEP-2000;	2000US-0236370P;
PR	29-SEP-2000;	2000US-0236371P;
PR	02-OCT-2000;	2000US-0236802P;
PR	02-OCT-2000;	2000US-0237037P;
PR	02-OCT-2000;	2000US-0237038P;
PR	02-OCT-2000;	2000US-0237039P;
PR	02-OCT-2000;	2000US-0237040P;
PR	13-OCT-2000;	2000US-0239935P;
PR	13-OCT-2000;	2000US-0239937P;
PR	20-OCT-2000;	2000US-0240960P;
PR	20-OCT-2000;	2000US-0241785P;
PR	20-OCT-2000;	2000US-0241786P;
PR	20-OCT-2000;	2000US-0241787P;
PR	20-OCT-2000;	2000US-0241808P;
PR	20-OCT-2000;	2000US-0241809P;
PR	20-OCT-2000;	2000US-0241826P;

PR	20-OCT-2000;	2000US-0242221P.
PR	01-NOV-2000;	2000US-0244617P.
PR	08-NOV-2000;	2000US-0246475P.
PR	08-NOV-2000;	2000US-0246475P.
PR	08-NOV-2000;	2000US-0246476P.
PR	08-NOV-2000;	2000US-0246477P.
PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246603P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246611P.
PR	08-NOV-2000;	2000US-0246613P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249264P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0249279P.
PR	17-NOV-2000;	2000US-0249300P.
PR	01-DEC-2000;	2000US-0249301P.
PR	01-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251160P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0256719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	11-DEC-2000;	2000US-0251990P.
PR	05-JAN-2001;	2000US-0254097P.
XX		
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX	Rosen CA, Barash SC, Ruben	
PI	WPI; 2001-541565/60.	
XX		
DR		
XX		
PT	Nucleic acids encoding 3224 h	
PT	useful for preventing, diagno	
PT	and metastases.	
XX		
XX		
PS	Disclosure; SEQ ID NO 8137; 1	
XX		
CC	The invention relates to novel	
CC	(AB014678-AB0138001) useful fo	
CC	medical conditions e.g. by pro	
CC	isolated from a range of huma	
CC	The nucleic acids, proteins	
CC	the diagnosis, treatment and	
CC	ovarian cancer and other can	
CC	breast, gastrointestinal tract	
CC	disorders e.g. Addison's dise	
CC	anaemia, autoimmune thyroidit	

CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 494 BP; 99 A; 124 C; 121 G; 150 T; 0 U; 0 Other;
SQ Query Match 4.7%; Score 47; DB 5; Length 494;
Best Local Similarity 100.0%; Pred. No. 7.1e-09;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 TCCCAAGTCTGGATTACAGCATGAGCCACCGCGCGGCCCC 115
Db 259 TCCCAAGTCTGGATTACAGCATGAGCCACCGCGCGGCCCC 305
RESULT 11
AAK71311
ID AAK71311 standard; DNA; 9888 BP.
XX AAK71311;
AC AAK71311;
XX 06-NOV-2001 (first entry)
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:26123.
DE Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW Cytostatic; Gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
XX WO200157182-A2.
XX 09-AUG-2001.
XX 17-JAN-2001; 2001WO-US0001354.
XX 31-JAN-2000; 2000US-0179065P.
PR 04-FEB-2000; 2000US-0180628P.
PR 24-FEB-2000; 2000US-0184654P.
PR 02-MAR-2000; 2000US-0186350P.
PR 16-MAR-2000; 2000US-0189874P.
PR 17-MAR-2000; 2000US-0190076P.
PR 18-APR-2000; 2000US-0198123P.
PR 19-MAY-2000; 2000US-0205515P.
PR 07-JUN-2000; 2000US-0209467P.
PR 28-JUN-2000; 2000US-0214886P.
PR 30-JUN-2000; 2000US-0215135P.
PR 07-JUL-2000; 2000US-0216647P.
PR 07-JUL-2000; 2000US-0216880P.
PR 11-JUL-2000; 2000US-0217487P.
PR 11-JUL-2000; 2000US-0217496P.
PR 14-JUL-2000; 2000US-0218290P.
PR 26-JUL-2000; 2000US-0220963P.
PR 26-JUL-2000; 2000US-0220964P.
PR 14-AUG-2000; 2000US-0224518P.
PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 18-AUG-2000; 2000US-0226279P.
PR 22-AUG-2000; 2000US-0226681P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234957P.
PR 25-SEP-2000; 2000US-0234988P.
PR 26-SEP-2000; 2000US-0235842P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 17-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.

XX OS Homo sapiens.
XX PN W0200159063-A2.
XX XX
XX PD 16-AUG-2001.
XX PF
XX PF 17-JAN-2001; 2001WO-US001334.
XX PF
XX PF 31-JAN-2000; 2000US-0179065P.
XX PF 04-FEB-2000; 2000US-0180628P.
XX PF 24-FEB-2000; 2000US-0184654P.
XX PF 02-MAR-2000; 2000US-0186350P.
XX PF 16-MAR-2000; 2000US-0189674P.
XX PF 17-MAR-2000; 2000US-0190076P.
XX PF 18-APR-2000; 2000US-0198123P.
XX PF 19-MAY-2000; 2000US-0205151P.
XX PF 07-JUN-2000; 2000US-0209467P.
XX PF 28-JUN-2000; 2000US-0214886P.
XX PF 30-JUN-2000; 2000US-0215135P.
XX PF 07-JUL-2000; 2000US-0216647P.
XX PF 07-JUL-2000; 2000US-0216880P.
XX PF 11-JUL-2000; 2000US-0217487P.
XX PF 11-JUL-2000; 2000US-0217496P.
XX PF 14-JUL-2000; 2000US-0218290P.
XX PF 26-JUL-2000; 2000US-0220963P.
XX PF 26-JUL-2000; 2000US-0220964P.
XX PF 14-AUG-2000; 2000US-0224518P.
XX PF 14-AUG-2000; 2000US-0224519P.
XX PF 14-AUG-2000; 2000US-0225213P.
XX PF 14-AUG-2000; 2000US-0225214P.
XX PF 14-AUG-2000; 2000US-0225266P.
XX PF 14-AUG-2000; 2000US-0225267P.
XX PF 14-AUG-2000; 2000US-0225268P.
XX PF 14-AUG-2000; 2000US-0225270P.
XX PF 14-AUG-2000; 2000US-0225447P.
XX PF 14-AUG-2000; 2000US-0225757P.
XX PF 14-AUG-2000; 2000US-0225758P.
XX PF 14-AUG-2000; 2000US-0225759P.
XX PF 18-AUG-2000; 2000US-0226279P.
XX PF 22-AUG-2000; 2000US-0226681P.
XX PF 22-AUG-2000; 2000US-0226686P.
XX PF 22-AUG-2000; 2000US-0227182P.
XX PF 23-AUG-2000; 2000US-0227009P.
XX PF 30-AUG-2000; 2000US-0228924P.
XX PF 01-SEP-2000; 2000US-0229287P.
XX PF 01-SEP-2000; 2000US-0229343P.
XX PF 01-SEP-2000; 2000US-0229344P.
XX PF 01-SEP-2000; 2000US-0229345P.
XX PF 05-SEP-2000; 2000US-0229509P.
XX PF 05-SEP-2000; 2000US-0229513P.
XX PF 06-SEP-2000; 2000US-0230437P.
XX PF 06-SEP-2000; 2000US-0230438P.
XX PF 08-SEP-2000; 2000US-0231242P.
XX PF 08-SEP-2000; 2000US-0231243P.
XX PF 08-SEP-2000; 2000US-0231244P.
XX PF 08-SEP-2000; 2000US-0231413P.
XX PF 08-SEP-2000; 2000US-0231414P.
XX PF 08-SEP-2000; 2000US-0232080P.
XX PF 08-SEP-2000; 2000US-0232081P.
XX PF 12-SEP-2000; 2000US-0231968P.
XX PF 14-SEP-2000; 2000US-0232397P.
XX PF 14-SEP-2000; 2000US-0232398P.
XX PF 14-SEP-2000; 2000US-0232399P.
XX PF 14-SEP-2000; 2000US-0232400P.
XX PF 14-SEP-2000; 2000US-0232401P.
XX PF 14-SEP-2000; 2000US-0233063P.
XX PF 14-SEP-2000; 2000US-0233064P.
XX PF 14-SEP-2000; 2000US-0233065P.
XX PF 21-SEP-2000; 2000US-0234223P.
XX PF 21-SEP-2000; 2000US-0234274P.
XX PF 25-SEP-2000; 2000US-0234997P.
XX PF 25-SEP-2000; 2000US-0234998P.
XX PF 26-SEP-2000; 2000US-0235484P.
XX PF 27-SEP-2000; 2000US-0235834P.
XX PF 27-SEP-2000; 2000US-0235836P.
XX PF 29-SEP-2000; 2000US-0236327P.
XX PF 29-SEP-2000; 2000US-0236367P.
XX PF 29-SEP-2000; 2000US-0236368P.
XX PF 29-SEP-2000; 2000US-0236369P.
XX PF 29-SEP-2000; 2000US-0236370P.
XX PF 02-OCT-2000; 2000US-0236802P.
XX PF 02-OCT-2000; 2000US-0237037P.
XX PF 02-OCT-2000; 2000US-0237038P.
XX PF 02-OCT-2000; 2000US-0237039P.
XX PF 13-OCT-2000; 2000US-0237040P.
XX PF 13-OCT-2000; 2000US-0239935P.
XX PF 13-OCT-2000; 2000US-0239937P.
XX PF 20-OCT-2000; 2000US-0240960P.
XX PF 20-OCT-2000; 2000US-0241785P.
XX PF 20-OCT-2000; 2000US-0241786P.
XX PF 20-OCT-2000; 2000US-0241787P.
XX PF 20-OCT-2000; 2000US-0241808P.
XX PF 20-OCT-2000; 2000US-0241809P.
XX PF 20-OCT-2000; 2000US-0241826P.
XX PF 20-OCT-2000; 2000US-0242221P.
XX PF 01-NOV-2000; 2000US-0244617P.
XX PF 08-NOV-2000; 2000US-0246474P.
XX PF 08-NOV-2000; 2000US-0246475P.
XX PF 08-NOV-2000; 2000US-0246476P.
XX PF 08-NOV-2000; 2000US-0246477P.
XX PF 08-NOV-2000; 2000US-0246478P.
XX PF 08-NOV-2000; 2000US-0246523P.
XX PF 08-NOV-2000; 2000US-0246524P.
XX PF 08-NOV-2000; 2000US-0246525P.
XX PF 08-NOV-2000; 2000US-0246526P.
XX PF 08-NOV-2000; 2000US-0246527P.
XX PF 08-NOV-2000; 2000US-0246528P.
XX PF 08-NOV-2000; 2000US-0246532P.
XX PF 08-NOV-2000; 2000US-0246609P.
XX PF 08-NOV-2000; 2000US-0246610P.
XX PF 08-NOV-2000; 2000US-0246611P.
XX PF 08-NOV-2000; 2000US-0246613P.
XX PF 17-NOV-2000; 2000US-0249207P.
XX PF 17-NOV-2000; 2000US-0249208P.
XX PF 17-NOV-2000; 2000US-0249209P.
XX PF 17-NOV-2000; 2000US-0249210P.
XX PF 17-NOV-2000; 2000US-0249211P.
XX PF 17-NOV-2000; 2000US-0249212P.
XX PF 17-NOV-2000; 2000US-0249213P.
XX PF 17-NOV-2000; 2000US-0249214P.
XX PF 17-NOV-2000; 2000US-0249215P.
XX PF 17-NOV-2000; 2000US-0249216P.
XX PF 17-NOV-2000; 2000US-0249217P.
XX PF 17-NOV-2000; 2000US-0249218P.
XX PF 17-NOV-2000; 2000US-0249244P.
XX PF 17-NOV-2000; 2000US-0249245P.
XX PF 17-NOV-2000; 2000US-0249264P.
XX PF 17-NOV-2000; 2000US-0249265P.
XX PF 17-NOV-2000; 2000US-0249297P.
XX PF 17-NOV-2000; 2000US-0249299P.
XX PF 17-NOV-2000; 2000US-0249300P.
XX PF 01-DEC-2000; 2000US-0250391P.
XX PF 01-DEC-2000; 2000US-0251160P.
XX PF 05-DEC-2000; 2000US-0251030P.
XX PF 05-DEC-2000; 2000US-0251988P.
XX PF 05-DEC-2000; 2000US-0256719P.
XX PF 06-DEC-2000; 2000US-0251479P.
XX PF 08-DEC-2000; 2000US-0251856P.
XX PF 08-DEC-2000; 2000US-0251869P.
XX PF 08-DEC-2000; 2000US-0251869P.
XX PF 08-DEC-2000; 2000US-0251989P.
XX PF 08-DEC-2000; 2000US-0251990P.
XX PF 11-DEC-2000; 2000US-0254097P.
XX PF 05-JAN-2001; 2001US-0259679P.
XX XX

PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PT useful for preventing, diagnosing and/or treating nervous system cancers
PT and metastases.
XX
XX Disclosure; SEQ ID NO 8136; 1701pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABAI1004-ABAI21534) and proteins
CC (ABAI4678-ABAI8001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer, and other cancers of the adrenal gland, bone, bone marrow,
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC infectious diseases such as viral, bacterial, fungal and parasitic
CC infections. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX SQ Sequence 13189 BP; 3249 A; 2881 C; 3117 G; 3942 T; 0 U; 0 Other;
SQ
Query Match 4.7%; Score 47; DB 5; Length 13189;
Best Local Similarity 100.0%; Pred. No. 4.2e-09;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 TCCCAAGTCTGGGATTACAGCATGAGCCACCGCGCGCGCCCC 115
DB 8116 TCCCAAGTCTGGGATTACAGCATGAGCCACCGCGCGCCCC 8162
RESULT 14
ADC86998/c
ID ADC86998 standard; DNA; 43419 BP.
XX
XX AC ADC86998;
XX
XX DT 01-JAN-2004 (first entry)
XX
XX DE Human GPCR gene SEQ ID NO:1451.
XX
XX KW ds; gene; human; GPCR;
XX KW guanosine triphosphate-binding protein coupled receptor; gene therapy.
XX
XX OS Homo sapiens.
XX
XX PN EP1270724-A2.
XX
XX PD 02-JAN-2003.
XX
XX PF 18-JUN-2002; 2002BP-00013517.
XX
XX PR 18-JUN-2001; 2001JP-00245789.
XX
XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
XX (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
XX
XX PI Suwa M, Asai K, Akiyama Y, Aburatani H;
XX
XX WPI; 2003-315783/31.
XX
XX DR P-PSDB; ADC86999.
XX
XX PT New polynucleotide, useful for preparing a composition for treating a

PT patient in need of increased or suppressed activity or expression of the
PT guanosine triphosphate-binding protein coupled receptor.
XX
XX PS Claim 1; SEQ ID NO 1451; 28pp; English.
XX
XX CC The invention relates to a novel polynucleotide encoding a guanosine
CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
CC the invention may have a use in gene therapy. The polynucleotide and
CC polypeptide are useful for preparing a composition for treating a patient
CC in need of increased or suppressed activity or expression of the
CC guanosine triphosphate-binding protein coupled receptor. The
CC polynucleotide sequences shown in ADC85548-ADC87616 encode GPCR's of the
CC invention.
XX
XX SQ Sequence 43419 BP; 11054 A; 10220 C; 10281 G; 11764 T; 0 U; 100 Other;
SQ
Query Match 4.7%; Score 47; DB 9; Length 43419;
Best Local Similarity 100.0%; Pred. No. 3.4e-09;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 52 ATCCGCCCGCTCAGCTTCCCAAAGTCTGGGATTACAGCATGAGC 98
DB 31644 ATCCGCCCGCTCAGCTTCCCAAAGTCTGGGATTACAGCATGAGC 31598
RESULT 15
ADC27000
ID ADC27000 standard; DNA; 63626 BP.
XX
XX AC ADC27000;
XX
XX DT 19-DEC-2003 (first entry)
XX
XX DE Human carcinoma associated (CA) genomic DNA PRDM11.
XX
XX KW Carcinoma associated gene; CA gene; PR domain protein 11; PRDM11;
XX KW carcinoma associated protein; CAP; carcinoma; cancer; adenocarcinoma;
XX KW breast cancer; inflammatory carcinoma; Paget's disease; aging;
XX KW neurodegenerative disease; cytostatic; osteopathic; neurotropic;
XX KW neuroprotective; human; ds.
XX
XX OS Homo sapiens.
XX
XX PN US2003087252-A1.
XX
XX PD 08-MAY-2003.
XX
XX PF 20-MAR-2002; 2002US-00105637.
XX
XX PR 22-DEC-2000; 2000US-00747377.
XX
XX PR 02-MAR-2001; 2001US-00798586.
XX
XX PR 20-DEC-2001; 2001US-00034650.
XX
XX FA (MORR/) MORRIS D W.
XX FA (ENGE/) ENGELHARD E K.
XX
XX PI Morris DW, Engelhard EK;
XX
XX DR WPI; 2003-730063/69.
XX
XX PT Novel recombinant carcinoma-associated gene, PRDM11 useful for diagnosing
XX PT cancer, and for screening for agents for treating cancers and
XX PT neurodegenerative diseases.
XX
XX PS Claim 1; Page 36-45; 49pp; English.
XX
XX CC The present invention relates to novel carcinoma associated (CA) nucleic
XX CC acid sequences from both mouse and human. In particular the invention
XX CC discloses a recombinant nucleic acid comprising a fully defined PR domain
XX CC protein 11 (PRDM11 - a carcinoma-associated gene) nucleotide sequence.
XX CC The sequences of the invention are useful for identifying compounds that
XX CC modulate the activity of a carcinoma associated protein (CAP) which are
XX CC potentially useful in treating carcinomas. Among the diseases and

CC conditions that may be diagnosed or treated according to the invention
 CC are cancers such as adenocarcinoma, breast cancer, inflammatory
 CC carcinoma, Paget's disease, etc, aging, and neurodegenerative diseases.
 CC The present sequence represents a CA nucleic acid sequence of the
 CC invention.

XX

SQ Sequence 63626 BP; 15622 A; 15444 C; 15753 G; 16795 T; 0 U; 12 Other;

Query Match 4.7%; Score 47; DB 9; Length 63626;

Best Local Similarity 100.0%; Pred. No. 3.2e-09;

Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 66 GCTTCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGCGCC 112

Db 18670 GCTTCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGCGCC 18716

Search completed: July 10, 2004, 13:27:05

Job time : 501 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2004, 13:04:20 ; Search time 3110 Seconds
(without alignments)
9505.967 Million cell updates/sec

Title: US-09-972-032-1

Perfect score: 990
Sequence: 1 ggaattgtctcgaggccaa.....aaaatgagcgccgaagt 990

Scoring table: OLIGO_NUC
Gapop_60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	562	56.8	1467	11	BC039500 Homo sapi
c 2	370	37.4	760	12	BM980506
c 3	343	34.6	691	13	BU620427
c 4	341	34.4	1201	13	EX406138

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

c 5	323	32.6	669	12	BM679577	BM679577 UI-E-EO0-
c 6	323	32.6	694	13	BU684594	BU684594 UI-CF-EN1
c 7	312	31.5	623	14	CD742581	CD742581 UI-H-FT2-
c 8	308	31.1	632	13	BU617331	BU617331 UI-H-DF0-
c 9	276	27.9	628	13	BQ447041	BQ447041 UI-H-EU1-
c 10	269	27.2	633	14	CA313142	CA313142 UI-CF-FN0
c 11	225	22.7	395	9	AA687318	AA687318 nv61e01.s
c 12	212	21.4	300	12	BM672504	BM672504 UI-E-CQ1-
c 13	211	21.3	462	12	BM687386	BM687386 UI-E-CQ1-
c 14	211	21.3	480	10	BF509528	BF509528 UI-H-B14-
c 15	210	21.2	367	10	BF002915	BF002915 7950h12.x
c 16	210	21.2	411	9	AI184808	AI184808 qd24g11.x
c 17	209	21.1	211	9	AI202243	AI202243 qs65d12.x
c 18	207	20.9	439	10	BE677246	BE677246 7d82e01.x
c 19	184	18.6	1084	12	BM806108	BM806108 AGENCOURT
c 20	103	10.4	282	10	BF364571	BF364571 PM1-NN108
c 21	57	5.8	659	29	AG160016	AG160016 Pan trogl
c 22	57	5.8	751	29	AG037002	AG037002 Pan trogl
c 23	55	5.6	688	29	AG118999	AG118999 Pan trogl
c 24	54	5.5	658	29	AG161334	AG161334 Pan trogl
c 25	54	5.5	740	28	AZ520059	AZ520059 RPC1-11-2
c 26	53	5.4	172	10	AW841136	AW841136 RC5-CN001
c 27	53	5.4	224	10	AW841118	AW841118 RC5-CN001
c 28	53	5.4	450	28	AQ331706	AQ331706 HS_5011_A
c 29	48	4.8	433	10	BE243207	BE243207 TCAAP2D05
c 30	48	4.8	715	29	AG153081	AG153081 Pan trogl
c 31	47	4.7	328	9	AI370057	AI370057 tb74g01.x
c 32	47	4.7	340	14	T02817	T02817 FB10B11 Pet
c 33	47	4.7	379	9	AI370094	AI370094 tb75c08.x
c 34	47	4.7	379	10	AW403177	AW403177 UI-HF-BK0
c 35	47	4.7	383	9	AI370074	AI370074 tb75a01.x
c 36	47	4.7	578	14	CA397714	CA397714 CS94f12.y
c 37	47	4.7	680	29	AG124506	AG124506 Pan trogl
c 38	46	4.6	155	13	BU571107	BU571107 AGENCOURT
c 39	46	4.6	386	9	AS514737	AS514737 nh71d11.s
c 40	46	4.6	407	10	BE152030	BE152030 QV1-HT031
c 41	46	4.6	414	28	B66825	B66825 CIT-HSP-201
c 42	46	4.6	423	14	H60216	H60216 YR38C07.r1
c 43	46	4.6	493	14	N92133	N92133 za23e12.r1
c 44	46	4.6	500	10	BF820071	BF820071 MRI-RT003
c 45	46	4.6	513	28	AQ235733	AQ235733 HS_2015_B

ALIGNMENTS

RESULT 1

BC039500

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

BC039500 Homo sapiens, clone IMAGE:5555626, mRNA.
1467 bp
linear HTC 04-MAR-2003

BC039500.1 GI:25058499

HTC.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1467)

Strausberg, R.

Direct Submission

Submitted (01-NOV-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.ncbi.nih.gov>

Contact: MGC help desk

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-sbhc.stanford.edu>

REMARK

COMMENT

Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 88 Row: g Column: 1
This clone has the following problem: retained intron.

FEATURES

source

```
1. 1467
  Location/Qualifiers
    organism="Homo sapiens"
    mol_type="mRNA"
    db_xref="taxon:9606"
    clone="IMAGE:5555626"
    tissue_type="Uterus, leiomyosarcoma"
    clone_lib="NIH MGC 71"
    lab_host="DH10B"
    note="vector: pCMV-SPORT6"
```

ORIGIN

Query Match 56.8%; Score 562; DB 11; Length 1467;

Best Local Similarity 99.7%; Pred. No. 0;

Matches 662; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy 34 CGATCTCTCGTACCTCGGTGATCCCGCGCTCAGCTTCCCAAGTGTGGATTACAGCA 93
Db |
Qy 504 CGATCTCTCGTACCTCGGTGATCCCGCGCTCAGCTTCCCAAGTGTGGATTACAGCA 563
Db |
Qy 94 TGAGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 153
Db |
Qy 564 TGAGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 623
Qy 154 TGGCGCGCGTCCAGCTGCGCGCGTCACTGCGCGCGTCACTGCGCGCGTCACTGCGCG 213
Db |
Qy 624 TGGCGCGCGTCCAGCTGCGCGCGTCACTGCGCGCGTCACTGCGCGCGTCACTGCGCG 683
Qy 214 ACCGCGTCCGCTAAGCGCTGATGCTGCTGCTGATGATGATGATGATGATGATGATGAT 273
Db |
Qy 684 ACCGCGTCCGCTAAGCGCTGATGCTGCTGCTGATGATGATGATGATGATGATGATGAT 743
Qy 274 TGGCGTAGGGGAAGGGCAGAGGGCGGGAATGGGCGGAGGCGGCGGCGGCGGCGGCGGCG 333
Db |
Qy 744 TGGCGTAGGGGAAGGGCAGAGGGCGGGAATGGGCGGAGGCGGCGGCGGCGGCGGCGGCG 803
Qy 334 CGACTGCGCTCCCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 393
Db |
Qy 804 CGACTGCGCTCCCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 863
Qy 394 CGTCCGAGATCGTCCGCAACTGGCGGAGTGTGATGCGGCGGCGTGGCTAAGGCGCGTGGTT 453
Db |
Qy 454 GGTACGATGTCAGCGGAGTAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 513
Db |
Qy 924 GGTACGATGTCAGCGGAGTAAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 983
Qy 514 AGGTCCTGGAAGAGTGTATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 573
Db |
Qy 984 AGGTCCTGGAAGAGTGTATCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1043
Qy 574 GGTGCGTGGCGCTTAGGGCGAGCGCGGAGATACCCGCGGAGTGGTGGATGATATACC 633
Db |
Qy 1044 GTGTGGTGGCGCTTAGGGCGAGCGCGGAGACACACCCCGGAGTGGTGGATGATATACC 1103
Qy 634 GGAAGGGGTGATCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693
Db |
Qy 1104 GGAAGGGGTGATCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1163
Qy 694 GGCA 697
Db |
Qy 1164 GGCA 1167

```

RESULT 2

BM980506/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

BM980506
UI-CF-EN1-add-a-08-0-UI.s1
UI-CF-EN1-add-a-08-0-UI 3', mRNA sequence.
BM980506
BM980506.1 GI:19602038
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 760)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
889548
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES

source

```
1..760
  Location/Qualifiers
    organism="Homo sapiens"
    mol_type="mRNA"
    db_xref="taxon:9606"
    clones="UI-CF-EN1-add-a-08-0-UI"
    tissue_type="Primary Lung Cystic Fibrosis Epithelial
    Cells"
    dev_stage="Adult"
    lab_host="DH10B (Life Technologies) (T1 phage resistant)"
    clone_lib="UI-CF-EN1"
    notes="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
    modified polylinker; Site 1: EcoR I; Site 2: Not I;
    UI-CF-EN1 is a normalized cDNA library containing the
    following tissue(s): Primary Lung Cystic Fibrosis
    Epithelial Cells. The library was constructed according to
    Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
    1996. First strand cDNA synthesis was primed with an
    oligo-dT primer containing a Not I site. Double stranded
    cDNA was ligated to an EcoR I adaptor, digested with Not
    I, and cloned directionally into pT73-Pac vector. The
    oligonucleotide used to prime the synthesis of
    first-strand cDNA contains a library tag sequence that is
    located between the Not I site and the (dT)18 tail. The
    sequence tag for this library is CTGCTCAGGT.
    TAG TISSUE=Human Lung Epithelial Cell lines untreated LPS
    6hr to LPS 24h
    TAG LIB=UI-CF-EN1
    TAG_SEQ=CTGCTCAGGT"
```

ORIGIN

Query Match 37.4%; Score 370; DB 12; Length 760;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 710; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
Qy 240 GCTTCGCTGATGCACATTGGACCGGCTCTGGACTGGGTAGGGAGGAGGGCGG 299
Db |
Qy 734 GCTTCGCTGATGCACATTGGACCGGCTCTGGACTGGGTAGGGAGGAGGGCGG 675
Db |

```

QY 300 GAATTGGGCGGAGGGCCAGGCGCTCGCGGACCCCGACTGCGCCTCCCGGTGGCCCGCGCA 359
Db 674 GAATTGGGCGGAGGGCCAGGCGCTCGCGGACCCCGANTGCGCTCCCGGTGGCCCGCGCA 615
QY 360 GCGCCTCCCGGTGGCCCTCGAGTGCAGGTCTTACCGTCCGAGATCGTCCGCAACTGGCGG 419
Db 614 GCGCCTCCCGGTGGCCCTCGAGTGCAGGTCTTACCGTCCGAGATCGTCCGCAACTGGCGG 555
QY 420 AGCTGTGCATGGGCGGTGGCTAAAGCGCGTGGTTGGTTAGATGGCCAGCGGGAATTAA 479
Db 554 AGCTGTGCATGGGCGGTGGCTAAAGCGCGTGGTTGGTTAGATGGCCAGCGGGAATTAA 495
QY 480 GTGTGTCTCTGAAGACATGGACATAGTCTCGAGGGTCCCTGGAGAGTATCCCGGCC 539
Db 494 GTGTGTCTCTGAAGACATGGACATAGTCTCGAGGGTCCCTGGAGAGTATCCCGGCC 435
QY 540 CCACCATCAATGGCGCTTAGGCTTAGGAAGCGGGTGTGGGTGGGCGCTTAGGGCGAGGC 599
Db 434 CCACCATCAATGGCGCTTAGGCTTAGGAAGCGGGTGTGGGTGGGCGCTTAGGGCGAGGC 375
QY 600 GCAGACATACCCGAGTGTGGATGTATACCGCAAGGGGTGGATCGAACCCCGCA 659
Db 374 GCAGACATACCCGAGTGTGGATGTATACCGCAAGGGGTGGATCGAACCCCGCA 315
QY 660 AGACACTGGAAGGCTGTGCGCTGAGGAGGCGCGGCA-ATCCAGTGTGCTGTGGCTTT 718
Db 314 AGACACTGGAAGGCTGTGCGCTGAGGAGGCGCGGCAATCCAGTGTGCTGTGGCTTT 255
QY 719 ACAGGAAGAGCTCCACCTTCT-TGGAGTGTGCAGATGCGATCTAGTGTGTCCACCGCA 777
Db 254 ACAGGAAGAGCTCCACCTTCT-TGGAGTGTGCAGATGCGATCTAGTGTGTCCACCGCA 195
QY 778 TGGGAGCTGCGGCGGAGAGTCTGCCCGAGTACAAAGCTGATTTGGACCTGGGCGCT 837
Db 194 TGGGAGCTGCGGCGGAGAGTCTGCCCGAGTACAAAGCTGATTTGGACCTGGGCGCT 135
QY 838 CTGACATTCCTGATTCCTGCTGTGATCTCCAGCAAGTCTGTCCCGTGGCTGCCCT 897
Db 134 CTGACATTCCTGATTCCTGCTGTGATCTCCAGCAAGTCTGTCCCGTGGCTGCCCT 75
QY 898 CATCCACTCTCATCTCTGCTTCAGAGTAAATTCAGATCTGTGGTGC 951
Db 74 CATCCACTCTCATCTCTGCTTCAGAGTAAATTCAGATCTGTGGTGC 21

RESULT 3
BU620427/c
LOCUS
DEFINITION
  UI-H-FL1-bfw-n-19-0-UI.s1 NCI CGAP FL1 Homo sapiens cDNA clone
  UI-H-FL1-bfw-n-19-0-UI 3', mRNA sequence.
ACCESSION
  BU620427
VERSION
  1 (bases 1 to 691)
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
  Unpublished (1997)
  Contact: Robert Strausberg, Ph.D.
  Email: cgabs-remail.nih.gov
  Tissue Procurement: James Martin
  CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Clone distribution information can be obtained
  from Dr. M. Bento Soares, bento-soares@uiowa.edu
  Seq primer: M13 FORWARD
  POLYA=Yes.

```

```

FEATURES
  source
  Location/Qualifiers
    1..691
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="UI-H-FL1-bfw-n-19-0-UI"
      /tissue_type="Cell lines"
      /dev_stage="Adult"
      /lab_host="DH10B (Life Technologies)"
      /clone_lib="NCI CGAP FL1"
      /note="Organ: Chondrosarcoma; Vector: pT73-Pac
      (Pharmacia) with a modified polylinker; Site 1: Ecor I;
      Site 2: Not I; NCI CGAP FL1 is a normalized cDNA library
      derived from a pool of mRNA obtained from 4 cell lines
      from grade III chondrosarcoma tissues. The library was
      constructed according to Bonaldo, Lemmon and Soares,
      Genome Research, 6:791-806, 1996. First strand cDNA
      synthesis was primed with an oligo-dT primer containing a
      Not I site. Double stranded cDNA was ligated to an EcoR I
      adaptor, digested with Not I, and cloned directionally
      into pT73-Pac vector. The oligonucleotide used to prime
      the synthesis of first-strand cDNA contains a library tag
      sequence that is located between the Not I site and the
      (dT)18 tail. The sequence tag for this library is
      GAGTCCGTG. The cell lines were provided by Dr. James
      Martin from the University of Iowa.
      TAG TISSUE=Human Chondrosarcoma Grade 3 cell line mix
      TAG_L1B=UI-H-FL1
      TAG_SEQ=GAGTCCGTG"

ORIGIN
  Query Match      34.6%; Score 343; DB 13; Length 691;
  Best Local Similarity 99.7%; Pred. No. 0;
  Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 304 TGGGCGGCGAGGGCCAGGCGCTCGCGGACCCCGACTGCGCCTCCCGGTGGCCCGCGCGC 363
Db 667 TGGGCGGCGAGGGCCAGGCGCTCGCGGACCCCGACTGCGCCTCCCGGTGGCCCGCGCGC 608
QY 364 CTCGCCGTGGCCCTCGAGTGCAGGTCTTACCGTCCGAGATCGTCCGCAACTGGGCGAGCT 423
Db 607 CTCGCCGTGGCCCTCGAGTGCAGGTCTTACCGTCCGAGATCGTCCGCAACTGGGCGAGCT 548
QY 424 GTGCATGGGCGGTGGCTAAAGCGCGGTGGTTGGTTACGATTGGCCAGCGGGACTTAAAGTGT 483
Db 547 GTGCATGGGCGGTGGCTAAAGCGCGGTGGTTGGTTACGATTGGCCAGCGGGACTTAAAGTGT 488
QY 484 TGCTCTCTGAAGAGCATGCACATTAGTCTGGAGGCTCTGGAAGAGTGTATCCCGCCCGCAC 543
Db 487 TGCTCTCTGAAGAGCATGCACATTAGTCTGGAGGCTCTGGAAGAGTGTATCCCGCCCGCAC 428
QY 544 CATCAATGGGCGCTTAGGTCTAGGAAGCGGGTGGGTGGGCGCTTAGGGCGAGCGCGAG 603
Db 427 CATCAATGGGCGCTTAGGTCTAGGAAGCGGGTGGGTGGGCGCTTAGGGCGAGCGCGAG 368
QY 604 ACATACCCCGAAGTGGTTGGAATTGATACCGAAGGGCTGGATCGAACCCCGCAAGAC 663
Db 367 ACACACCCCGAAGTGGTTGGAATTGATACCGAAGGGCTGGATCGAACCCCGCAAGAC 308
QY 664 ACTGGAAGGCTGTGTGGCTGAGGAGCGGCCCGCGCA 697
Db 307 ACTGGAAGGCTGTGTGGCTGAGGAGCGGCCCGCGCA 274

RESULT 4
BX406138/c
LOCUS
DEFINITION
  BX406138 Homo sapiens PENTAL LIVER Homo sapiens cDNA clone
  CS0DM010YD19 3-PRIME, mRNA sequence.
ACCESSION
  BX406138
VERSION
  BX406138.1 GI:30648317
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)

```



```

Db      522  ACGATTGGCCAGCGGACTTAAGTGTGTCTCTGAAGAGCATGACATTAGTCTGGAGG 463
Qy      518  TCTTGAAGAGTATCCCGCCGCCACCATCAAAATGGCGTTAGTCTAGAAAGCGGTGT 577
Db      462  TCTTGAAGAGTATCCCGCCGCCACCATCAAAATGGCGTTAGTCTAGAAAGCGGTGT 403
Qy      578  GGGTGGGGCTTAGGGGAGCGCAGACATACCCCGAAGTGGTTGGATTGTATACCGCAA 637
Db      402  GGGTGGGGCTTAGGGGAGCGCAGACATACCCCGAAGTGGTTGTATACCGCAA 343
Qy      638  GGGGCTGGATCGAACCCGCCCAAGACACTGGAGGCTGTGTGCTGAGAGGGCCCGGCA 697
Db      342  GGGGCTGGATCGAACCCGCCCAAGACACTGGAGGCTGTGTGCTGAGAGGGCCCGGCA 283
Qy      698  -ATCCAGTGTGTCTGGGGCTTTACAGAAAGAGCTCCACCTTCT-TGGAGTGGCAGATG 755
Db      282  GATCCAGTGTGTCTGGGGCTTTACAGAAAGAGCTCCACCTTCTTGGAGTGGCAGATG 223
Qy      756  CGATCTAGGTGTGTCCACCGATGGGAGTGGCGGGCGGCGCAGATGCTCCCCAGTACAA 815
Db      222  CGATCTAGGTGTGTCCACCGATGGGAGTGGCGGGCGGCGCAGATGCTCCCCAGTACAA 163
Qy      816  AGCTGATTTGGACTGGGGCTCTGGACTTCCCTGATTTCTCTCTTGCATCTCCAGCAA 875
Db      162  AGCTGATTTGGACTGGGGCTCTGGACTTCCCTGATTTCTCTCTTGCATCTCCAGCAA 103
Qy      876  GTCTGTCTCCCGTTGGCTTCCATCCACTCTCTCACTTCTCTGCTTCAGAGTAAATTT 935
Db      102  GTCTGTCTCCCGTTGGCTTCCATCCACTCTCTCACTTCTCTGCTTCAGAGTAAATTT 43
Qy      936  GCAAGATCTGTGGTC 951
Db      42  GCAAGATCTGTGGTC 27

```

RESULT 6

```

BU684594/c
LOCUS      BU684594          694 bp      mRNA      linear      EST 07-OCT-2002
DEFINITION UI-CF-EN1-act-a-22-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
            UI-CF-EN1-act-a-22-0-UI 3', mRNA sequence.
ACCESSION  BU684594
VERSION    BU684594.1 GI:23537704
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 694)
            Bonaldo M.F., Lennon G. and Soares M.B.
            Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL    Genome Res. 6 (9), 791-806 (1996)
MEDLINE    97044477
PUBMED     8889548
COMMENT    Contact: McCray, PB
            University of Iowa
            2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
            Tel: 319 356 4866
            Fax: 319 356 7171
            Email: paul-mccray@uiowa.edu
            Tissue Procurement: Dr. M. J. Welsh, University of Iowa
            CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com) or from Open Biosystems
            (www.openbiosystems.com).
            Seq primer: M13 FORWARD
            POLYA-Yes.
            Location/Qualifiers
            1..694
            /organism="Homo sapiens"

```

FEATURES

```

source
1..694
/organism="Homo sapiens"
936  GCAAGATCTGTGGTC 951

```

```

/mol_type="mRNA"
/db_xref="taxon:9606"
/clones="UI-CF-EN1-act-a-22-0-UI"
/tissue_type="Primary Lung Cystic Fibrosis Epithelial
Cells"
/dev_stages="Adult"
/lab_host="DH10B (Life Technologies) (T1 phase resistant)"
/clone_lib="UI-CF-EN1"
/notes="Organ: Lung; Vector: p773-Pac (Pharmacia) with a
modified polylinker; Site_1: Sc0r 1; Site_2: Not 1;
UI-CF-EN1 is a normalized cDNA library containing the
following tissue(s): Primary Lung Cystic Fibrosis
Epithelial Cells. The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was ligated to an EcoR I adaptor, digested with Not
I, and cloned directionally into p773-Pac vector. The
oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tag for this library is CTGCTCAGGT.
TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS
6hr to LPS 24h
TAG_LIB=UI-CF-EN1
TAG_SEQ=CTGCTCAGGT"

```

ORIGIN

```

Query Match      32.6%; Score 323; DB 13; Length 694;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 613; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
Qy      338  TCGCGCTCCCGGTGGCCCGCAGCGCCCTCCCGGTGGCCCTCGAGTGCAGGCTTACCGTC 397
Db      647  TCGCGCTCCCGGTGGCCCGCAGCGCCCTCCCGGTGGCCCTCGAGTGCAGGCTTACCGTC 588
Qy      338  CGAGATCGTCGCAACTGGCGAGCTGTGCATGGGGCGTGGCTAAGGCCGTGGTTGGTT 457
Db      587  CGAGATCGTCGCAACTGGCGAGCTGTGCATGGGGCGTGGCTAAGGCCGTGGTTGGTT 528
Qy      458  ACGATTGGCCAGCGGACTTAAAGTGTCTCTGAAGAGCATGGACATTAGTCTGGAGGG 517
Db      527  ACGATTGGCCAGCGGACTTAAAGTGTCTCTGAAGAGCATGGACATTAGTCTGGAGGG 468
Qy      518  TCTCGAAGAGTGATCCCGCCCGCCACCATCAATAGGCGCTTAGGTCTAGGAAGCGGGTGT 577
Db      467  TCTCGAAGAGTGATCCCGCCCGCCACCATCAATAGGCGCTTAGGTCTAGGAAGCGGGTGT 408
Qy      578  GGCTGGGGCTTAGGGCGAGCGCAGACATACCCCGAGTGGTTGGATTGTATACCGCAA 637
Db      407  GGCTGGGGCTTAGGGCGAGCGCGAGACACACCCCGAAGTGGTTGGATTGTATACCGCAA 348
Qy      638  GGGGCTGGATCGAACCCCGCCAAAGACACTGGAAAGCTGTGTGGCTGAGGAGGGCCCGGCA 697
Db      347  GGGGCTGGATCGAACCCCGCCAAAGACACTGGAAAGCTGTGTGGCTGAGGAGGGCCCGGCA 288
Qy      698  -ATCCAGTGTGTGTGGGCTTTACAGAAAGAGCTCCACCTTCT-TGAGTGTGCAGATG 755
Db      287  GATCCAGTGTGTGTGGGCTTTACAGAAAGAGCTCCACCTTCTCTGAGTGTGCAGATG 228
Qy      756  CGATCTAGGTGTGTCCACCGATGGAGCTCCGGCGGGCGAGATGCTGCCCGCAGTACAA 815
Db      227  CGATCTAGGTGTGTCCACCGATGGAGCTCCGGCGGGCGAGATGCTGCCCGCAGTACAA 168
Qy      816  AGCTGATTTGGACTGGGGCTCTGGACTTCCCTGATTTCTCTGCTTCAGTCTCCAGCAA 875
Db      167  AGCTGATTTGGACTGGGGCTCTGGACTTCCCTGATTTCTCTGCTTCAGTCTCCAGCAA 108
Qy      876  GTCTGTCTCCCGTTGGCTTCCATCCACTCTCTCACTTCTCTGCTTCAGAGTAAATTT 935
Db      107  GTCTGTCTCCCGTTGGCTTCCATCCACTCTCTCACTTCTCTGCTTCAGAGTAAATTT 48
Qy      936  GCAAGATCTGTGGTC 951

```

```

Db      47 GCAGATCTGTGGTGC 32
|||||
RESULT 7
CD742581/c
LOCUS      623 bp mRNA linear EST 26-JUN-2003
DEFINITION UI-H-PT2-bj1-i-10-0-UI.s1 NCI CGAP_FT2 Homo sapiens cDNA clone
            UI-H-PT2-bj1-i-10-0-UI 3', mRNA sequence.
ACCESSION  CD742581
VERSION     CD742581.1 GI:32293431
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 623)
            NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapsb-remail.nih.gov
            Tissue Procurement: Dr. Gary W. Hunninghake, U of I
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Distribution information can be found at
            http://genome.uiowa.edu/distribution/cgap.html
            Seq primer: M13 FORWARD
            POLYA=yes.
FEATURES             Location/Qualifiers
     source           1..623
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-PT2-bj1-i-10-0-UI"
                     /tissue_type="Aveolar Macrophage"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI CGAP FT2"
                     /notes="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site 1: EcoR I; Site 2: Not I;
                     NCI-CCAP_FT2 is a subtracted cDNA library constructed from
                     a pool of 81 RNA samples from Alveolar Macrophages
                     challenged with different treatments. The library was
                     subtracted according to Bonaldo, Lennon and Soares, Genome
                     Research, 6:791-806, 1996. The tissue was provided by Dr.
                     Gary W. Hunninghake of the University of Iowa.
                     TAG_TISSUE=Human Lung Alveolar Macrophage
                     TAG_LIB=UI-H-PT2
                     TAG_SEQ=GGCCATGCGG"
ORIGIN
Query Match      31.5%; Score 312; DB 14; Length 623;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 602; Conservative 0; Mismatches 1; Indels 2; Gaps 2;
QY      343 CTCGGTGGCCCGCAGCGCCCTCCCGTGGCCCTCGAGTGCAGGCTTACCGTCCGAGA 402
Db      623 CTCGGTGGCCCGCAGCGCCCTCCCGTGGCCCTCGAGTGCAGGCTTACCGTCCGAGA 564
QY      403 TCGTCGCACTGGCGAGCTGTCATGGGCGTGGCTAAGCGCGGTTTGGTTACGAT 462
Db      563 TCGTCGCACTGGCGAGCTGTCATGGGCGTGGCTAAGCGCGGTTTGGTTACGAT 504
QY      463 TGGCCAGCGGACTTAAGTGTCTCTGAAGAGCATGGACATTAAGTCTGGAGGGTCTCTG 522
Db      503 TGGCCAGCGGACTTAAGTGTCTCTGAAGAGCATGGACATTAAGTCTGGAGGGTCTCTG 444
QY      523 GAGAGTGATCCCGCCACCATCAATGCGGCTTAGCTTAGGAGCGGGTGGGTG 582
Db      443 GAAGAGTGATCCCGCCACCATCAATGCGGCTTAGCTTAGGAGCGGGTGGGTG 384

```

```

QY      583 GGGCTTAGGGCGGAGCGGCAGACATACCCGAACTGGTTGGATTGTATACCGCAAGGGGC 642
Db      383 GGGCTTAGGGCGGAGCGGCAGACACACCCGAACTGGTTGGATTGTATACCGCAAGGGGC 324
QY      643 TGGATCCGAACCCCAAGACACTGGAAGGCTGTGTGCTGAGGAGGCGCCGGCA-ATCC 701
Db      323 TGGATCCGAACCCCAAGACACTGGAAGGCTGTGTGCTGAGGAGGCGCCGGCAGATCC 264
QY      702 AGTGTGCTGTGGCTTTACAGGAAAGAGCTCCACTTCT-TGGAGTGTGAGATGCCATC 760
Db      263 AGTGTGCTGTGGCTTTACAGGAAAGAGCTCCACTTCTCTGAGAGTGTGAGATGCCATC 204
QY      761 TAGGTGTCTCCACCCGATGGGAGCTGGGCGCGGCGAGATGCTGCCCCAGTACAAAGCTG 820
Db      203 TAGGTGTCTCCACCCGATGGGAGCTGGGCGCGGCGAGATGCTGCCCCAGTACAAAGCTG 144
QY      821 ATTTGGACCTGGGCGCTCTGGACTTCCCTGATTCCTGTGCTTGCATCTCCAGCAAAAGTCT 880
Db      143 ATTTGGACCTGGGCGCTCTGGACTTCCCTGATTCCTGTGCTTGCATCTCCAGCAAAAGTCT 84
QY      881 GTCCGTTGGCTGCTTCATCCACTCTCTCACTTCTCTGCTTCAGAGTAAATTTGCAAG 940
Db      83 GTCCGTTGGCTGCTTCATCCACTCTCTCACTTCTCTGCTTCAGAGTAAATTTGCAAG 24
QY      941 ATCTG 945
Db      23 ATCTG 19

```

```

RESULT 8
BU617331/c
LOCUS      632 bp mRNA linear EST 23-SEP-2002
DEFINITION UI-H-DF0-bep-n-09-0-UI.s1 NCI CGAP_DF0 Homo sapiens cDNA clone
            UI-H-DF0-bep-n-09-0-UI 3', mRNA sequence.
ACCESSION  BU617331
VERSION     BU617331.1 GI:23283539
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 632)
            NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapsb-remail.nih.gov
            Tissue Procurement: Dr. Jose Mercuende
            cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Clone distribution information can be obtained
            from Dr. M. Bento Soares, bento-soares@uiowa.edu
            Seq primer: M13 FORWARD
            POLYA=yes.
FEATURES             Location/Qualifiers
     source           1..632
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-H-DF0-bep-n-09-0-UI"
                     /tissue_type="Subchondral Bone"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="NCI CGAP_DF0"
                     /notes="Organ: Bone; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site 1: EcoR I; Site 2: Not I;
                     NCI CGAP_DF0 is a cDNA library containing the following
                     tissue(s): Subchondral Bone. The library was constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed

```


McCrack Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccrack@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=Yes.
Location/Qualifiers
1. .633
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-FNO-23-0-UI"
/tissue_type="Human Lung Epithelial cells"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-FNO"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site 1: EcoR I; Site 2: Not I;
UI-CF-FNO is a subcloned cDNA library derived from two
normalized Human lung epithelial cell libraries (EN1 and
DUI). The library was subcloned according to according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. For additional information, contact:
bento-soares@uiowa.edu
TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS
6hr to LPS 24h
TAG LIB=UI-CF-FNO
TAG_SEQ=CTGCTCAGGT"

FEATURES
source
Query Match 27.2%; Score 269; DB 14; Length 633;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 609; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
QY 341 GCCTCCGGTGGCCCGCAGCGCTCCCGTGGCCCTGGAGTCAGCTTACCGCCGA 400
DB 633 GCCTCCGGTGGCCCGCAGCGCTCCCGTGGCCCTGGAGTCAGCTTACCGCCGA 574
QY 401 GATCGTCCGCACTGGGGAGCTGTGATGGGGCGTGAAGCCCGTGGTTTACG 460
DB 573 GATCGTCCGCACTGGGGAGCTGTGATGGGGCGTGAAGCCCGTGGTTTACG 514
QY 461 ATTGCCAGCGGAGCTTAAGTGTGTCTCTGAAGAGCATGACATTAAGTCGAGGGTCC 520
DB 513 ATTGCCAGCGGAGCTTAAGTGTGTCTCTGAAGAGCATGACATTAAGTCGAGGGTCC 454
QY 521 TGGAGAGTGTATCCCGCCACCATCAATGGCGCTTAGTCTAGGAAGCGGTGTGGG 580
DB 453 TGGAGAGTGTATCCCGCCACCATCAATGGCGCTTAGTCTAGGAAGCGGTGTGGG 394
QY 581 TGGGCGCTTAGGGGAGCGCAGACATACCCCGAAGTGTGTATACCGCAAGGG 640
DB 393 TGGGCGCTTAGGGGAGCGCAGACATACCCCGAAGTGTGTATACCGCAAGGG 334
QY 641 GCTGGATCGAACCCCGCAGACATGGAAGCTGTGTGGCTAGGAGGCGCGGCA-AT 699
DB 333 GCTGGATCGAACCCCGCAGACATGGAAGCTGTGTGGCTAGGAGGCGCGGCAAT 274
QY 700 CCAAGTGTGTGGGCTTTACAGGAAGAGCTCCACCTTC-TTGGAGTGTGCAATCGGA 758
DB 273 CCAAGTGTGTGGGCTTTACAGGAAGAGCTCCACCTTC-TTGGAGTGTGCAATCGGA 214
QY 759 TCTAGGTGTGCCACCGATGGAGCTCGGGGCGGAGATGCTGCCCGCAGTACAAGC 818
DB 213 TCTAGGTGTGCCACCGATGGAGCTCGGGGCGGAGATGCTGCCCGCAGTACAAGC 154

QY 819 TGATTGGACCTGGGGCTCTGGACTTCCCTGATTCCTCTGCTTGCATCTCCAGCAAGTC 878
DB 153 TGATTGGACCTGGGGCTCTGGACTTCCCTGATTCCTCTGCTTGCATCTCCAGCAAGTC 94
QY 879 CTGTCCTGGCTGGCTGGCTTCATCCACTCTCTCACTTCTGCTTGCATCTCCAGCAAGTC 938
DB 93 CTGTCCTGGCTGGCTGGCTTCATCCACTCTCTCACTTCTGCTTGCATCTCCAGCAAGTC 34
QY 939 AGACTCTGGTGC 951
DB 33 AGACTCTGGTGC 21
RESULT 11
AA687318/c
LOCUS AA687318
DEFINITION nvlsl01.sl NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1234296 3',
mRNA sequence.
ACCESSION AA687318
VERSION AA687318.1 GI:2675509
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 955)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
NATIONAL Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
unknown library type
Insert Length: 966 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 361.
FEATURES
source
1. .395
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:1234296"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_GCB1"
/note="Vector: pT73-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I-Oligo (dT) primer
15'-GTGTACCAATCTGAAGTGGAGCGCGCTCATTTTCTTTTCTTTT-3'
1. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

Query Match 22.7%; Score 225; DB 9; Length 395;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 741 TGGAGTGTGCAATGCGATCTAGGTGTGTCCACCGAGTGGAGTGGCGGCGGAGAT 800
DB 225 TGGAGTGTGCAATGCGATCTAGGTGTGTCCACCGAGTGGAGTGGCGGCGGAGAT 166
QY 801 GCTGCCCGCAGTACAAGCTGATTTGGAGCTGGGGCTCTGGACTTCCTGATTTCTGCT 860
DB 165 GCTGCCCGCAGTACAAGCTGATTTGGAGCTGGGGCTCTGGACTTCCTGATTTCTGCT 106
ORIGIN
Query Match 22.7%; Score 225; DB 9; Length 395;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 861 TGCATCTCCAGCAAGTCTGTCCGTTGGTGGCTTCATCCACTCTCTCACTTCTCTGC 920
Db 105 TGCATCTCCAGCAAGTCTGTCCGTTGGTGGCTTCATCCACTCTCTCACTTCTCTGC 46
QY 921 CTTTCAGAGTAAATTCGAAGATCTGTGGTGCACAAAAA 965
Db 45 CTTTCAGAGTAAATTCGAAGATCTGTGGTGCACAAAAA 1

RESULT 12
BM672504/c
LOCUS
DEFINITION
  UI-E-CQ1-ada-e-09-0-UI.s1 UI-E-CQ1 Homo sapiens cDNA clone
BM672504
ACCESSION
  BM672504.1 GI:118982402
VERSION
  EST.
KEYWORDS
  Homo sapiens (human)
SOURCE
  Homo sapiens
ORGANISM
  Homo sapiens
REFERENCE
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 300)
  Bernaldo,M.F., Lennon,G. and Soares,M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
  Genome Res. 6 (9), 791-806 (1996)
JOURNAL
  9704477
MEDLINE
  8889548
PUBMED
  Contact: Soares, MB
  Coordinated Laboratory for Computational Genomics
  University of Iowa
  375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
  Tel: 319 335 8250
  Fax: 319 335 9565
  Email: bento-soares@uiowa.edu
  Tissue Procurement: Dr. Gregg Hageman
  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
  DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Researchers may obtain clones from Research
  Genetics (www.resgen.com).
  The following repetitive elements were found in this cDNA
  sequence: 1-21, >AT rich#Low_complexity (matched complement)
  Seq primer: M13 Forward
  POLYA=Yes.
FEATURES
  source
    1..300
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="UI-E-CQ1-ada-e-09-0-UI"
    /tissue_type="optic nerve"
    /dev_stage="adult"
    /lab_host="DHI0B (Life Technologies) (T1 phage resistant)"
    /clone_lib="UI-E-CQ1"
    /notes="Organ: eye; Vector: p77T3-Pac (Pharmacia) with a
    modified polylinker; Site 1: EcoR I; Site 2: Not I;
    UI-E-CQ1 is a normalized cDNA library containing the
    following tissue(s): optic nerve. The library was
    constructed according to Bernaldo, Lennon and Soares,
    Genome Research, 6:791-806, 1996. First strand cDNA
    synthesis was primed with an oligo-dT primer containing a
    Not I site. Double stranded cDNA was ligated to an EcoR I
    adaptor, digested with Not I, and cloned directionally
    into p77T3-Pac vector. The oligonucleotide used to prime
    the synthesis of first-strand cDNA contains a library tag
    sequence that is located between the Not I site and the
    (dT)18 tail. The sequence tag for this library is
    CCAATTAAGT. This library was created for the program, Gene
    Discovery in the Visual System, supported by National Eye
    Institute (NEI).
    TAG TISSUE=human optic nerve
    TAG_LIB=UI-E-CQ1

```

TAG_SEQ=CCATTAAGTC

ORIGIN

```

Query Match 21.4%; Score 212; DB 12; Length 300;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 212; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 740 TTGGAGTGTGCAGATCGCATCTAGGTGTGTCCACCCGATGGAGCTGGGGCGGGCAGA 799
Db 233 TTGGAGTGTGCAGATCGCATCTAGGTGTGTCCACCCGATGGAGCTGGGGCGGGCAGA 174
QY 800 TGCTGCCCCAGTACAAAGCTGATTTGGACCTGGGGCCTCTGGACTTCCCTGATTTCTGC 859
Db 173 TGCTGCCCCAGTACAAAGCTGATTTGGACCTGGGGCCTCTGGACTTCCCTGATTTCTGC 114
QY 860 TTGCATCTCCAGCAAGTCTGTCCGTTGGTGGCTTCATCCACTCTCTCACTTCTCTG 919
Db 113 TTGCATCTCCAGCAAGTCTGTCCGTTGGTGGCTTCATCCACTCTCTCACTTCTCTG 54
QY 920 CTTTCAGAGTAAATTCGAAGATCTGTGGTGC 951
Db 53 CTTTCAGAGTAAATTCGAAGATCTGTGGTGC 22

```

RESULT 13

```

BM687366
LOCUS
DEFINITION
  UI-E-CQ1-ada-e-09-0-UI.r1 UI-E-CQ1 Homo sapiens cDNA clone
BM687366
ACCESSION
  UI-E-CQ1-ada-e-09-0-UI 5', mRNA sequence.
VERSION
  BM687366.1 GI:19000624
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 452)
  Ronaldo,M.F., Lennon,G. and Soares,M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
  Genome Res. 6 (9), 791-806 (1996)
JOURNAL
  9704477
MEDLINE
  8889548
PUBMED
  Contact: Soares, MB
  Coordinated Laboratory for Computational Genomics
  University of Iowa
  375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
  Tel: 319 335 8250
  Fax: 319 335 9565
  Email: bento-soares@uiowa.edu
  Tissue Procurement: Dr. Gregg Hageman
  cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
  DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Researchers may obtain clones from Research
  Genetics (www.resgen.com).
  Seq primer: M13 Reverse
  Location/Qualifiers
    1..462
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="UI-E-CQ1-ada-e-09-0-UI"
    /tissue_type="optic nerve"
    /dev_stage="adult"
    /lab_host="DHI0B (Life Technologies) (T1 phage resistant)"
    /notes="Organ: eye; Vector: p77T3-Pac (Pharmacia) with a
    modified polylinker; Site 1: EcoR I; Site 2: Not I;
    UI-E-CQ1 is a normalized cDNA library containing the
    following tissue(s): optic nerve. The library was
    constructed according to Bernaldo, Lennon and Soares,
    Genome Research, 6:791-806, 1996. First strand cDNA
    synthesis was primed with an oligo-dT primer containing a
    Not I site. Double stranded cDNA was ligated to an EcoR I
    adaptor, digested with Not I, and cloned directionally
    into p77T3-Pac vector. The oligonucleotide used to prime
    the synthesis of first-strand cDNA contains a library tag
    sequence that is located between the Not I site and the
    (dT)18 tail. The sequence tag for this library is
    CCAATTAAGT. This library was created for the program, Gene
    Discovery in the Visual System, supported by National Eye
    Institute (NEI).
    TAG TISSUE=human optic nerve
    TAG_LIB=UI-E-CQ1

```

FEATURES

source

synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p7733-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CCATTAGTG. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

ORIGIN

Query Match 21.3%; Score 211; DB 12; Length 462;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 211; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 TGGAGTGTGCAGATGGCATCTAGTGTGTCCACCCGATGGAGCTCGGGCCGGGCAGAT 800
|||||
Db 252 TGGAGTGTGCAGATGGCATCTAGTGTGTCCACCCGATGGAGCTCGGGCCGGGCAGAT 311
|||||

QY 801 GCTGCCCCAGTACAAAGCTGATTGGACCTGGGGCTCTGGACTTCCCTGATCTCTGCT 860
|||||

Db 312 GCTGCCCCAGTACAAAGCTGATTGGACCTGGGGCTCTGGACTTCCCTGATCTCTGCT 371
|||||

QY 861 TGCATCTCCAGCAAGTCTCCCTGGCTGGCTGCTTCATCCACTCTCTCACTTCTCTGC 920
|||||

Db 372 TGCATCTCCAGCAAGTCTCCCTGGCTGGCTGCTTCATCCACTCTCTCACTTCTCTGC 431
|||||

QY 921 CTTTCAGAGTAAATTCGAAGATCTGTGGTGC 951
|||||

Db 432 CTTTCAGAGTAAATTCGAAGATCTGTGGTGC 462
|||||

RESULT 14

BF509528/c
LOCUS
DEFINITION
UI-H-B14-aos-f-10-0-UI.s1 NCI_CGAP_Sub8 Homo sapiens cDNA clone
IMAGE:3086731 3', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BF509528.1 GI:11592826
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 480)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA library Preparation: M.B. Soares Lab Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward
POLYA=Yes.

FEATURES

source

1. 480
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3086731"
/lab_hosts="DH10B (Life Technologies)"
/clone_lib="NCI_CGAP_Sub8"
/notes="Vector: p7733D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; NCI_CGAP_Sub8 is a subtracted library derived from NCI_CGAP_Sub5. The NCI_CGAP_Sub8 library had 2.5 million recombinants. A

single-stranded DNA preparation of NCI_CGAP_Subs was used as a tracer in a subtractive hybridization with a driver comprising: a pool of clones from NCI_CGAP_Sub5 (IMAGE clone ids 2732833-2737415, 3068040-3069191; 25% of the driver population), a pool of clones from NCI_CGAP_Sub4 (IMAGE clone ids 2723592-2729326; 25% of the driver population), NCI_CGAP_Sub6 (pool AIF-AJU, IMAGE ids 2728989-2733190; 25% of the driver population), and NCI_CGAP_Sub7 (IMAGE ids 3069192-3072238, 3081864-3084550; 25% of the driver population). Subtraction was performed as previously described (Bonaldo, Lennon & Soares (1996): Normalization and Subtraction: Two Approaches To Facilitate Gene Discovery. Genome Research 6, 791-806.

TAG_TISSUE=lung
TAG_LIB=NCI_CGAP_Lul9
TAG_SEQ=GACAGC"

ORIGIN

Query Match 21.3%; Score 211; DB 10; Length 480;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 211; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 741 TGGAGTGTGCAGATGGCATCTAGTGTGTCCACCCGATGGAGCTCGGGCCGGGCAGAT 800
Db 230 TGGAGTGTGCAGATGGCATCTAGTGTGTCCACCCGATGGAGCTCGGGCCGGGCAGAT 171
QY 801 GCTGCCCCAGTACAAAGCTGATTGGACCTGGGGCTCTGGACTTCCCTGATCTCTGCT 860
Db 170 GCTGCCCCAGTACAAAGCTGATTGGACCTGGGGCTCTGGACTTCCCTGATCTCTGCT 111
QY 861 TGCATCTCCAGCAAGTCTCCCTGGCTGGCTGCTTCATCCACTCTCTCACTTCTCTGC 920
Db 110 TGCATCTCCAGCAAGTCTCCCTGGCTGGCTGCTTCATCCACTCTCTCACTTCTCTGC 51
QY 921 CTTTCAGAGTAAATTCGAAGATCTGTGGTGC 951
Db 50 CTTTCAGAGTAAATTCGAAGATCTGTGGTGC 20

RESULT 15

BF002915/c
LOCUS
DEFINITION
795Ch12.x1 NCI_CGAP_Pr28 Homo sapiens cDNA clone IMAGE:3309959 3', mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BF002915
BF002915.1 GI:10703190
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 367)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco.

FEATURES

source

1..367
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

```
/clone="IMAGE:3309959"  
/sex="male"  
/dev_stage="adult"  
/lab_host="DH10B"  
/clone_lib="NCI_CGAP_Pr28"  
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)  
with a modified polylinker; Plasmid DNA from the  
normalized library NCI CGAP Pr22 was prepared, and ss  
circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (cloneids  
985608-986759, 1101192-1101959, and 1217928-1220615).  
Subtraction by Bento Soares and M. Fatima Bonaldo. "
```

ORIGIN

```
Query Match      21.2%; Score 210; DB 10; Length 367;  
Best Local Similarity 100.0%; Pred. NO. 0;  
Matches 210; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 741 TGGAGTGTGCAGATGCGATCTAGTGTGTCCACCCGATGGAGCTGGGGCCGGGCAGAT 800  
    |||||  
Db 214 TGGAGTGTGCAGATGCGATCTAGTGTGTCCACCCGATGGAGCTGGGGCCGGGCAGAT 155  
    |||||  
Qy 801 GCTGCCCCAGTACAAAGCTGATTGGACCTGGGGCCCTCTGGACTTCCTGATTCTCTGCT 860  
    |||||  
Db 154 GCTGCCCCAGTACAAAGCTGATTGGACCTGGGGCCCTCTGGACTTCCTGATTCTCTGCT 95  
    |||||  
Qy 861 TGATCTCCAGCAAGTCCCTGTCCGTTGCTTCCTTCATCCACTCTTCACATTCCTCTGC 920  
    |||||  
Db 94  TGATCTCCAGCAAGTCCCTGTCCGTTGCTTCCTTCATCCACTCTTCACATTCCTCTGC 35  
    |||||  
Qy 921 CTTCAGAGTAAATTCGAAGATCTGTGGTG 950  
    |||||  
Db 34  CTTCAGAGTAAATTCGAAGATCTGTGGTG 5
```

Search completed: July 10, 2004, 15:29:05
Job time : 3115 secs

This Page Blank (uspto)

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	579	58.5	182230	9	AC135050	AC135050 Homo sapi
2	508	51.3	195476	2	AC135044	AC135044 Homo sapi
C 3	79	8.0	96975	9	AC004975	AC004975 Homo sapi
C 4	66	6.7	129758	9	AC118269	AC118269 Homo sapi
C 5	66	6.7	138536	9	AC116914	AC116914 Homo sapi
C 6	66	6.7	140210	2	AC002993	AC002993 Homo sapi
7	66	6.7	209885	2	AC011189	AC011189 Homo sapi
8	63	6.4	143065	9	HS20208	AL031848 Human DNA
C 9	60	6.1	113872	9	HS57029	Z95116 Human DNA
C 10	59	6.0	44174	9	AC004645	AC004645 Homo sapi
C 11	59	6.0	45454	9	AF157412	AL157412 Human DNA
C 12	59	6.0	122029	9	AF212832	AF212832 Homo sapi
C 13	59	6.0	123291	9	AC104070	AC104070 Homo sapi
C 14	59	6.0	148963	2	BX571818	BX571818 Homo sapi
C 15	59	6.0	159619	2	AC012440	AC012440 Homo sapi
16	59	6.0	166843	9	AC090821	AC090821 Homo sapi
17	59	6.0	169003	2	AC026071	AC026071 Homo sapi
18	59	6.0	169893	2	AC021165	AC021165 Homo sapi
C 19	59	6.0	181558	2	AC068301	AC068301 Homo sapi
20	59	6.0	189207	9	AC103706	AC103706 Homo sapi
21	59	6.0	190814	9	AC006111	AC006111 Homo sapi
22	59	6.0	199776	9	AC011499	AC011499 Homo sapi
C 23	59	6.0	213056	9	AC103590	AC103590 Homo sapi
C 24	58	5.9	105574	2	AC013308	AC013308 Homo sapi
25	58	5.9	111388	9	AC016655	AC016655 Homo sapi
26	58	5.9	115487	2	AC027344	AC027344 Homo sapi
27	58	5.9	118447	2	AC093508	AC093508 Homo sapi
28	58	5.9	124752	9	AC010419	AC010419 Homo sapi
29	58	5.9	132117	9	AC092463	AC092463 Homo sapi
C 30	58	5.9	160956	2	AC013637	AC013637 Homo sapi
31	58	5.9	161144	9	AC074142	AC074142 Homo sapi
32	58	5.9	170000	2	AC004524	AC004524 Homo sapi
33	58	5.9	183003	9	AC106045	AC106045 Homo sapi
34	58	5.9	197087	9	AC009964	AC009964 Homo sapi
35	58	5.9	200000	2	AC004604	AC004604 Homo sapi
36	58	5.9	210000	2	AC004555	AC004555 Homo sapi
37	57	5.8	115663	9	AC093029	AC093029 Homo sapi
38	57	5.8	138063	9	AC092473	AC092473 Homo sapi
C 39	57	5.8	156831	2	AC074268	AC074268 Homo sapi
C 40	57	5.8	162104	9	AC026325	AC026325 Homo sapi
C 41	57	5.8	213546	2	AC136602	AC136602 Homo sapi
C 42	56	5.7	164028	9	AC004890	AC004890 Homo sapi
C 43	56	5.7	167679	9	AC007570	AC007570 Homo sapi
44	56	5.7	186829	9	BX640519	BX640519 Human DNA
45	56	5.7	270178	2	BX572623	BX572623 Homo sapi

ALIGNMENTS

RESULT 1
AC135050
LOCUS
AC135050
DEFINITION
Homo sapiens chromosome 16 clone RP11-196G11, complete sequence.
ACCESSION
AC135050
VERSION
AC135050.3
KEYWORDS
HTG.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 182230)
AUTHORS
DOE Joint Genome Institute, Stanford Human Genome Center and Los
Alamos National Laboratory.
TITLE
Direct Submission

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: July 10, 2004, 11:58:39 ; Search time 4183 Seconds
(without alignments)
10458.093 Million cell updates/sec

Title: US-09-972-032-1
Perfect score: 990
Sequence: 1 ggaatgttctcgaggccaa.....aaaatgagcgccgaagtt 990

Scoring table: OLIGO NUC
Gapop_60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pi.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pin.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

QY	673	CTGTGTGGCTGAGAGGGCCCGGCA-ATCCAGTGTGTGTGGCTTTTACAGAAAGAGCT	731
Db	98559	CTGTGTGGCTGAGAGGGCCCGGCAATCCAGTGTGTGTGGCTTTTACAGAAAGAGCT	98618
QY	732	CCACCTTCT-TGAGTGTGCAGATGCAGATCTAGGTGTGTCCACCCGATGGAGCTGGGG	790
Db	98619	CCACCTTCTCTGAGTGTGCAGATGCAGATCTAGGTGTGTCCACCCGATGGAGCTGGGG	98678
QY	791	CCGGGCAGATGCTGCCCGACATACAAAGCTGATTTGGACCTGGGGCTCTGGACATTCCTCTG	850
Db	98679	CCGGGCAGATGCTGCCCGACATACAAAGCTGATTTGGACCTGGGGCTCTGGACATTCCTCTG	98738
QY	851	ATTCTCTGCTTGCATCTCCAGCAAGTCTCTGCTCCGTTGGCTTGCCTTCATCATCTCTCTC	910
Db	98739	ATTCTCTGCTTGCATCTCCAGCAAGTCTCTGCTCCGTTGGCTTGCCTTCATCATCTCTCTC	98798
QY	911	ACTTCTCTGCTTCCAGATGATTAATTCGAAGATCTCTGGTGC	951
Db	98799	ACTTCTCTGCTTCCAGATGATTAATTCGAAGATCTCTGGTGC	98839
RESULT 2			
AC135044			
LOCUS	AC135044	Homo sapiens chromosome 16 clone CTD-2551B20, WORKING DRAFT	195476 bp DNA linear HTG 05-OCT-2002
DEFINITION	AC135044	SEQUENCE, 22 unordered pieces.	
ACCESSION	AC135044		
VERSION	AC135044.1	GI:23505535	
KEYWORDS	HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE			
AUTHORS	1. (Bases 1 to 195476)		
TITLE	Sequencing of Human Chromosome 16		
JOURNAL	Unpublished		
REFERENCE	2. (Bases 1 to 195476)		
AUTHORS	DOE Joint Genome Institute.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-OCT-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
COMMENT	-----Genome Center Center: Joint Genome Institute Center Code: JGI Web site: http://www.jgi.doe.gov		
Project Information			
Center Project Name:	809609		
Center clone name:	CITB-E1_2551B20		
Summary Statistics			
Consensus quality:	171229 bases at least Q40		
Consensus quality:	182638 bases at least Q30		
Consensus quality:	188095 bases at least Q20		
Estimated insert size:	170000; agarose-fp estimation		
Estimated insert size:	193376; sum-of-contigs estimation		
Quality coverage:	7.63 in Q20 bases; agarose-fp estimation		
Quality coverage:	6.7 in Q20 bases; sum-of-contigs estimation.		
* NOTE:	This is a 'working draft' sequence. It currently consists of 22 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.		
* 1	1202: contig of 1202 bp in length		
* 1303	1302: gap of unknown length		
* 1303	2468: contig of 1166 bp in length		
* 2469	2568: gap of unknown length		
* 2569	4077: contig of 1509 bp in length		


```
* 4078 4177: gap of unknown length
* 4178 5772: contig of 1595 bp in length
* 5773 5872: gap of unknown length
* 5873 7722: contig of 1850 bp in length
* 7723 7822: gap of unknown length
* 7823 8958: contig of 1136 bp in length
* 8959 9059: gap of unknown length
* 9059 11052: contig of 1993 bp in length
* 11052 11152: gap of unknown length
* 11152 13642: contig of 2490 bp in length
* 13642 13742: gap of unknown length
* 13742 17738: contig of 3997 bp in length
* 17738 17838: gap of unknown length
* 17839 22030: contig of 4192 bp in length
* 22031 22131: gap of unknown length
* 22131 27003: contig of 4872 bp in length
* 27003 32819: gap of unknown length
* 32819 32919: contig of 5717 bp in length
* 32919 39003: gap of unknown length
* 39003 39103: contig of 6084 bp in length
* 39103 48370: gap of unknown length
* 48370 48470: contig of 9267 bp in length
* 48470 59721: gap of unknown length
* 59721 59821: contig of 11251 bp in length
* 59821 72304: gap of unknown length
* 72304 72404: contig of 12483 bp in length
* 72404 85349: gap of unknown length
* 85349 85449: contig of 12945 bp in length
* 85449 101872: gap of unknown length
* 101872 101972: contig of 16423 bp in length
* 101972 117076: gap of unknown length
* 117076 117176: contig of 15104 bp in length
* 117176 138973: gap of unknown length
* 138973 139074: contig of 21797 bp in length
* 139074 150336: gap of unknown length
* 150336 150437: gap of unknown length
* 150437 195476: contig of 45040 bp in length.
FEATURES             Location/Qualifiers
     source            1..195476
                        /organism="Homo sapiens"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:9606"
                        /chromosome="16"
                        /clone="CTD-2551B20"
                        /clone_lib="Caltech human BAC library D"
ORIGIN
Query Match          51.3%; Score 508; DB 2; Length 195476;
Best Local Similarity 99.8%; Pred. No. 3.2e-272;
Matches 558; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 139 TGCATTGTGGGAGACGGGCTCGCTAAGCGTGGATGTGGCTGATGCACATTG 258
DB 106367 TGCATTGTGGGAGACGGGCTCGCTAAGCGTGGATGTGGCTGATGCACATTG 198
QY 199 TGCATTGTGGGAGACGGGCTCGCTAAGCGTGGATGTGGCTGATGCACATTG 258
DB 106427 TGCATTGTGGGAGACGGGCTCGCTAAGCGTGGATGTGGCTGATGCACATTG 106486
QY 259 GACCGGGCTCTGGACTGGCTAGGGAGAGGCGAGGCGGCGAATTGGGCGCCGAGGGCCA 318
DB 106487 GACCGGGCTCTGGACTGGCTAGGGAGAGGCGAGGCGGCGAATTGGGCGCCGAGGGCCA 106546
QY 319 GGCTTCGCGACCCCGGCTCGGCTCCCGGTGGCGCCGCGAGGCGCTCCCGGTGGCGCTG 378
DB 106547 GGCTTCGCGACCCCGGCTCGGCTCCCGGTGGCGCCGCGAGGCGCTCCCGGTGGCGCTG 106606
QY 379 GAGTGCAGGCTTTACCGTCCGAGATCTCCGCAACTGGGCGAGCTGCGATGGGGCGGTGG 438
DB 106607 GAGTGCAGGCTTTACCGTCCGAGATCTCCGCAACTGGGCGAGCTGCGATGGGGCGGTGG 106666
QY 439 CTAAGCGCGGTGGTTGGTTAGCATTTGCCAGCGGACTTAAGTGTGCTCTCTGAGAGCA 498
```

```
Db 106667 CTAAGCGCGGTGGTTGGTTAGCATTTGCCAGCGGACTTAAGTGTGCTCTGAGAGCA 106726
QY 499 TGGACATTAGTCTGGAGGGTCTTGGAGAGTGTATCCCGCCCGCCACCATCAAAATGGCGCTT 558
DB 106727 TGGACATTAGTCTGGAGGGTCTTGGAGAGTGTATCCCGCCCGCCACCATCAAAATGGCGCTT 106786
QY 559 AGGTCTAGGAAGCGGGTGTGGGTGGGGCTTAGGGGAGGCGGAGACATACCCCGAAGTG 618
DB 106787 AGGTCTAGGAAGCGGGTGTGGGTGGGGCTTAGGGGAGGCGGAGACATACCCCGAAGTG 106846
QY 619 GTTGGATTGTATACCGCAAGGGGCTGGATCGAACCCCGCCCGCCCGCCCGAAGCTGTGT 678
DB 106847 GTTGGATTGTATACCGCAAGGGGCTGGATCGAACCCCGCCCGCCCGAAGCTGTGT 106906
QY 679 GGCTGAGAGGGGCCCGGCA 697
DB 106907 GGCTGAGAGGGGCCCGGCA 106925
RESULT 3
AC004975/c
LOCUS                AC004975                96975 bp    DNA    linear    PRI 21-DEC-1999
DEFINITION            Homo sapiens PAC clone RPS-1142J19 from 7q35-q36, complete
                        sequence.
ACCESSION              AC004975
VERSION                AC004975.2    GI:4753261
KEYWORDS                HTG.
SOURCE                  Homo sapiens (human)
ORGANISM                Homo sapiens
                        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE               1 (bases 1 to 96975)
                        Sulston,J.E. and Waterston,R.
                        Toward a complete human genome sequence
                        Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE                99063792
PUBMED                9847074
REFERENCE               2 (bases 1 to 96975)
                        Leonard,S. and Strommatt,C.
                        The sequence of Homo sapiens PAC clone RPS-1142J19
AUTHORS                Unpublished
JOURNAL                3 (bases 1 to 96975)
                        Waterston,R.H.
REFERENCE              Direct Submission
AUTHORS                Submitted (12-JUN-1998) Genome Sequencing Center, Washington
                        University School of Medicine, 4444 Forest Park Parkway, St. Louis,
                        MO 63108, USA
JOURNAL                MO 63108, USA
REFERENCE              4 (bases 1 to 96975)
                        Waterston,R.H.
AUTHORS                Direct Submission
JOURNAL                Submitted (27-AUG-1999) Department of Genetics, Washington
                        University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE              6 (bases 1 to 96975)
                        Waterston,R.
AUTHORS                Direct Submission
JOURNAL                Submitted (21-DEC-1999) Department of Genetics, Washington
                        University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
                        On May 5, 1999 this sequence version replaced gi:3213022.
COMMENT                ----- Genome Center
                        Center: Washington University Genome Sequencing Center
                        Center code: WUGSC
                        Web site: http://genome.wustl.edu/gsc
                        Contact: sapiens@wustl.wustl.edu
                        ----- Summary Statistics
                        -----
                        Center project name: H_DJ1142J19
                        -----
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GPB/CH7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

This clone was derived from human PAC library RPI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc.

VECTOR: pCYPAC2
(http://www.iesg.org)

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the right is RP4-814D15. Actual start of this clone is at base position 1 of RP5-1142J19; actual end is at base position 96975 of RP5-1142J19.

A transposon was identified in the cloning vector for RP5-1142J19.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/STB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
This clone was derived from human PAC library RP01-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.
VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP4-814D15. Actual start of this clone is at base position 1 of RP5-1142J19; actual end is at base position 96975 of RP5-1142J19.

A transposon was identified in the cloning vector for RP5-1142J19.

FEATURES	Location/Qualifiers
source	1..96975
	/organism="Homo sapiens"
	/mol_type="Genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7q35-q36"
	/clone="RP5-1142J19"
	/clone_lib="RPC1-5"
repeat_region	532..843
	/rpt_family="Alu"
repeat_region	853..883
	/rpt_family="AT-rich"
repeat_region	886..923
	/rpt_family="Alu"
repeat_region	930..1194
	/rpt_family="Alu"
repeat_region	2043..2080
	/rpt_family="L1"
repeat_region	2081..2382
	/rpt_family="Alu"
repeat_region	2383..2466
	/rpt_family="L1"
repeat_region	2467..2780
	/rpt_family="Alu"
repeat_region	2781..2796
	/rpt_family="L1"
repeat_region	2882..2910
	/rpt_family="(CAAAA)n"
repeat_region	2911..3190
	/rpt_family="Alu"
repeat_region	3325..3883
	/rpt_family="Retroviral"

O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viell,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
Submitted (21-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 129758)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Collangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., DeAllarano,K., Dewar,K., Diaz,J.S., Dodge,S., Fato,S., Ferreira,P., Fitzgerald,M., FitzHugh,W., Gage,D., Galagan,J., Gardaya,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kellis,C., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viell,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission
Submitted (24-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 21, 2002 this sequence version replaced gi:21392520.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RX/RepeatMasker.html>

----- Genome Center -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L26202
Center clone name: 2383_D_6

Location/Qualifiers
1..129758
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="CTD-2383D6"
/clone.lib="CITD1 Human BAC"
complement(2..224)
/rpt_family="AluJo"
complement(365..413)
/rpt_family="L2"
415..546
/rpt_family="AluSq/x"
549..844
/rpt_family="AluSp"
851..1158
/rpt_family="AluXx"
complement(1177..1928)
/rpt_family="L2"
5198..5315
/rpt_family="FLAM A"

```

repeat_region complement(5370..5680)
unsure /rpt_family="AluSg"
5466..5484 /note="single clone coverage"
repeat_region complement(6042..6349)
repeat_region /rpt_family="AluY"
complement(6856..7014)
repeat_region /rpt_family="AluJb"
complement(7020..7326)
repeat_region /rpt_family="AluSx"
7404..7719
/rpt_family="AluSx"
/rpt_family="AluSx"
complement(8263..8548)
repeat_region /rpt_family="AluSx"
complement(8550..8760)
/rpt_family="L1PAL0"
8828..9158
/rpt_family="AluSg1"
complement(9234..9436)
repeat_region /rpt_family="MER20"
complement(10017..10326)
/rpt_family="AluSx"
10476..10781
/rpt_family="AluSx"
complement(11225..11484)
/rpt_family="L2"
complement(11549..11657)
/rpt_family="MER105"
complement(11658..11969)
/rpt_family="AluSg"
complement(11970..12000)
/rpt_family="MER105"
complement(12001..12099)
/rpt_family="LTR65"
complement(12100..12401)
/rpt_family="AluSx"
complement(12402..12750)
/rpt_family="LTR65"
complement(12751..13057)
/rpt_family="AluY"
complement(13058..13208)
/rpt_family="LTR65"
complement(13213..13285)
/rpt_family="MER105"
complement(13708..13847)
/rpt_family="AluJb"
complement(13848..14151)
/rpt_family="AluSp"
complement(14152..14294)
/rpt_family="AluJb"
complement(14427..14567)
/rpt_family="MIR"
14568..14845
/rpt_family="AluSx"
14846..15018
/rpt_family="AluSg"
15360..15476
/rpt_family="L3"
complement(15478..15525)
/rpt_family="AluY"
complement(15588..15696)
/rpt_family="L2"
15697..16005
/rpt_family="AluSx"
complement(16006..16315)
/rpt_family="L2"
complement(16879..17153)
/rpt_family="AluSg"
17227..17421
/rpt_family="MER3"
17577..18116
/rpt_family="L2"
complement(19052..19360)

```

Query Match 6.7%; Score 66; DB 9; Length 129758;
 Best Local Similarity 100.0%; Pred. No. 3.5e-25;
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 CGATCTCTGACCTCGTATCCGCCCGCTCAGCTTCCCAAAGTCTGGGATTACAGGCA 93
 Db 95562 CGATCTCTGACCTCGTATCCGCCCGCTCAGCTTCCCAAAGTCTGGGATTACAGGCA 95503
 QY 94 TGAGCC 99
 Db 95502 TGAGCC 95497

RESULT 5
 AC116914/c 138536 bp DNA linear PRI 06-AUG-2002
 LOCUS Homo sapiens chromosome 17, clone CTD-319515, complete sequence.
 DEFINITION
 AC116914
 ACCESSION
 AC116914.5 GI:22038652
 VERSION
 HTG.
 KEYWORDS
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (Bases 1 to 138536)
 AUTHORS
 Birren,B., Nusbaum,C. and Lander,E.
 TITLE
 Homo sapiens chromosome 17, clone CTD-319515
 JOURNAL
 Unpublished
 REFERENCE
 2 (Bases 1 to 138536)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
 Bouckhalter,B., Brown,A., Camarata,J., Campoliano,A., Chang,J.,
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Fero,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gordon,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Horton,L., Hulme,W., Iliiev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
 Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
 Minova,I., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Trigglio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (Bases 1 to 138536)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
 Bouckhalter,B., Brown,A., Camarata,J., Campoliano,A., Chang,J.,
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Fero,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Gordon,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliiev,I.,
 Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
 Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
 Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
 Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
 Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
 Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
 Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
 Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,

Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainour, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL

Young, G.; Zainoun, J. B.; Zembek, D.; Zimmer, A. and Zou, M.
Direct Submission
Submitted (26-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE
AUTHORS

4 (bases 1 to 138536)
Birren, B., Nusbaum, C., Lander, B., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collins, A.,
Cook, A., Cooke, P., DeAtellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kama, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, N., Meldrum, J., Menelus, L., Mihova, T., Menga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schuppback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Scange-Thomann, N., Stojanovic, N., Talamas, J.,
Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliou, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL:

Zeng, D., Zinner, A. and Zou, M.
Direct Submission
Submitted (01-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
5 (bases 1 to 138536)

REFERENCE
AUTHORS

5 (bases 1 to 138536)
Barren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Birnba, N., Bateman, V., Bloom, T., Boguslavsky, L., Bouckhalter, B.,
Cannata, J., Chang, J., Chazaro, B., Choepel, Y., Collins, A.,
Comarata, J., Cooke, P., DeAtellano, K., Dewar, K., Diaz, J., S. Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galaagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kellis, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrim, J., Menius, L., Mihova, T., Menga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,
O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL:

Zemke, D., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 1, 2002 this sequence version replaced at:215911993.

COMMENT

On Aug 1, 2002 this sequence version replaced G1:21591993.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997).
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE JOURNAL

COMMENT

Submitted (17-FEB-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 4, 2001 this sequence version replaced g1:13959235.
All repeats were identified using RepeatMasker:

Smit, A. F. A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project information

Center project name: L487

Center clone name: 231_G_16

----- Summary Statistics

Sequencing vector: M13; M77815; 41% of reads

Sequencing vector: L08752; 4% of reads

Sequencing vector: Plasmid; n/a; 55% of reads

Chemistry: Dye-primer-amersham; 8% of reads

Chemistry: Dye-terminator Big Dye; 92% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 203261 bases at least Q40

Consensus quality: 205916 bases at least Q30

Consensus quality: 207122 bases at least Q20

Insert size: 177000; agarose-fp

Insert size: 208485; su

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 2927: contig of 2927 bp in length
2 2928 3027: gap of 100 bp
3 3028 4189: contig of 1162 bp in length
4 4190 4289: gap of 100 bp
5 4290 5459: contig of 1180 bp in length
6 5470 5569: gap of 100 bp
7 5570 6556: contig of 1087 bp in length
8 6557 6756: gap of 100 bp
9 6757 8150: contig of 1394 bp in length
10 8151 9645: contig of 1395 bp in length
11 9646 9745: gap of 100 bp
12 9746 11119: contig of 1374 bp in length
13 11120 11219: gap of 100 bp
14 11220 13069: contig of 1850 bp in length
15 13070 13169: gap of 100 bp
16 13170 15065: contig of 1896 bp in length
17 15066 15165: gap of 100 bp
18 15166 20167: contig of 5002 bp in length
19 20168 20267: gap of 100 bp
20 20268 29708: contig of 9341 bp in length
21 29709 45638: gap of 100 bp
22 45639 45738: contig of 15930 bp in length
23 45739 73550: contig of 27812 bp in length
24 73551 73650: gap of 100 bp
25 73651 138571: contig of 64921 bp in length
26 138572 138671: gap of 100 bp
27 138672 209885: contig of 71214 bp in length.

```

FEATURES Source

```

1. .209885
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="17"
   /map="17"

```

```

/misc_feature
/clone="RP11-231G16"
/clone_lib="RPC1-11 Human Male BAC"
1. .2927
/note="assembly_fragment"
clone_end:SP6
vector_side:left
3028. .4189
/note="assembly_fragment"
4290. .5469
/note="assembly_fragment"
5570. .6656
/note="assembly_fragment"
6757. .8150
/note="assembly_fragment"
8251. .9645
/note="assembly_fragment"
9746. .11119
/note="assembly_fragment"
11220. .13069
/note="assembly_fragment"
13170. .15065
/note="assembly_fragment"
15166. .20167
/note="assembly_fragment"
20268. .29608
/note="assembly_fragment"
29709. .45638
/note="assembly_fragment"
45739. .73550
/note="assembly_fragment"
73651. .138571
/note="assembly_fragment"
138672. .209885
/note="assembly_fragment"

```

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

ORIGIN

Query Match 6.7%; Score 66; DB 2; Length 209885;
Best Local Similarity 100.0%; Pred. No. 3.6e-25;
Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 34 CGATCTCTGACCTCGATCCGCCGCCCTCCAGCTTCCCAAGCTGGGATTACAGGCA 93

Db 181394 CGATCTCTGACCTCGATCCGCCGCCCTCCAGCTTCCCAAGCTGGGATTACAGGCA 181453

Oy 94 TGAGCC 99

Db 181454 TGAGCC 181459

RESULT 8

HS20208

LOCUS

DEFINITION

Human DNA sequence from clone RP1-20208 on chromosome 1p36.11-36.31

Contains the 5' part of a gene for a novel rat Espin Like protein

containing Ank repeats, the gene for the ortholog of rodent HES2

(Hairy and Enhancer of Split 2) and the 5' end of the gene for HBACH

(Brain Acyl-CoA Hydrolase (Acyl Coenzyme A Thioester Hydrolase, EC

3.1.2.2). Contains ESTs, GSSs and putative CpG islands, complete

sequence.

ACCESSION

AL031848

VERSION

AL031848.11 GI:4914512

KEYWORDS

HTG; Acyl Coenzyme A Thioester Hydrolase; Ank repeat; Brain

Acyl-CoA Hydrolase; Enhancer of Split; Espin; Hairy; HBACH; HES2.

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 143065)

Howden, P.

Direct Submission

Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,

Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk

COMMENT

On May 28, 1999 this sequence version replaced gi:4678811.

----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquery@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em.: EMBL; Sw.: SWISSPROT; Tr.: TREMBL; Wp.: WORMPEP; Information on the WORMPEP database can be found at

<http://www.sanger.ac.uk/projects/C.elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/Chrl>

RP1-20208 is from the library RPCT-1 constructed by the group of Pieter de Jong. For further details see

<http://www.chori.org/bacpac/home.htm>

VECTOR: pCYPAC2

This sequence is the entire insert of clone RP1-20208 The true left end of clone RP1-120232 is at 135606 in this sequence.

FEATURES

source

1..143065
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="RZPD:RPCIP704008202"
/db_xref="taxon:9606"
/chromosome="1"
/map="p36.11-36.31"
/clone="RP1-20208"
/clone_lib="RPCT-1"
7..125
/notes="MIR repeat: matches 13. .144 of consensus"
repeat_region
127..219
/notes="MER5A repeat: matches 14. .106 of consensus"
repeat_region
365..404
/notes="Alu repeat: matches 126. .165 of consensus"
repeat_region
405..634
/notes="AluSg/x repeat: matches 83. .309 of consensus"
misc_feature
complement(635..807)
/notes="match: GSS: Em:B42699"
misc_feature
complement(635..895)
/notes="match: GSS: Em:B47252"
gene
complement(732..27887)
/genes="dJ20208.1"

mRNA
complement(join(<732..995,1080..1225,3737..3935,4026..4187,6893..7164,8146..8347,11763..11894,12020..12202,12388..12574,24409..24602,27579..27887))
/genes="dJ20208.1"
/product="dJ20208.1 (novel rat Espin LIKE protein containing Ank repeats)"
/notes="match: CDSs: Em:U46007 Em:AL035288 Em:AF076856 Em:AF134401
match: ESTs: Em:H72801"
/evidence="not_experimental"
complement(join(<732..995,1080..1225,3737..3935,4026..4187,6893..7164,8146..8347,11763..11894,12020..12202,12388..12574,24409..24602,27579..27872))

CDS

/gene="dJ20208.1"
/note="match: proteins: Tr:Q63618 Tr:Q24341 Tr:O89019 Sw:Q01484 Tr:Q17343 Sw:Q01485 Tr:Q25338 Tr:Q92625"
/codon_start=1
/evidence="not_experimental"
/product="dJ20208.1 (novel rat Espin LIKE protein containing Ank repeats)"
/protein_id="CAB46197.1"
/db_xref="GI:5327035"
/translation="MALEQALQARQGLDVLSSLAAGLGPRLDPLDALPVHAA RAGKLCRLVVEALPAARARNATPAHDASATGHLACLQWLSSQGCRCVQKDN SGATVLAARFAGHPEVNVNLLHGGDPTAATDMGALPIHYAAAGDPFSLKLLVHE YPEVNAQTNGATPLYLACQEGHLEVTVYQVECCGAPAHADQMTFLHAAQVGH SPVVMVLSCTDVSQKDXGDAHMFASRGHTRVLSWLLHGGELISADLWGGTPL HDAENGELCCQILVNVGAELDRDRDGTAAADLDFNGHSHCTYLRVTENLSVEH RVLSRDPSSALEAKQPSDGMSSPNTTVSQPLNFDLSSPTSLSNYDSCSSSHSIKG QHPCGLSSARAADIOSYMDMLNPELGLPRGTIGKTPPPPPPPPPPPPPGCTQLPP PPGYPAPRPVGPQADIIYMQYKLRHVETALKEKELSSCDGHGDLRQDSSRKPR AFQKQSTGYRQLGRCPGCTLAARPGMAHSEEAALLPNNHVPNGCAADPRAARSLP PPPPPPPPLPEAASPPPPAPLPLESGAGCCQRRSSSTGSKFNMSPPTGDNSE LLAIRKAGSLKPTQSGKLTTFVSGIQGPAFQDPSLPSPALSPVRSPTPPAAGF QPLINGSLVPEVPTTPAPGVQLDVEALIPTHDEQGRPIPEWKQVNVKWLQKQEE EQRK"
778..1198
/notes="CpG island"
/evidence="not_experimental"
2896..3126
/notes="match: GSS: Em:AQ036085"
3347..3415
/notes="3 copies 23 mer 81% conserved"
3441..3507
/notes="MIR repeat: matches 57. .121 of consensus"
3702..5880
/notes="CpG island"
/evidence="not_experimental"
5573..5686
/notes="6 copies 19 mer 91% conserved"
6023..6423
/notes="L1M3e repeat: matches -370. .12 of consensus"
6999..7074
/notes="38 copies 2 mer gg 71% conserved"
7708..7741
/notes="17 copies 2 mer gt 82% conserved"
7709..7888
/notes="4 copies 45 mer 68% conserved"
9466..9591
/notes="MIR repeat: matches 31. .162 of consensus"
9640..9931
/notes="AluSg repeat: matches 1. .299 of consensus"
9940..10240
/notes="AluSc repeat: matches 1. .300 of consensus"
10241..10539
/notes="AluSc repeat: matches 1. .294 of consensus"
10556..10642
/notes="MIR repeat: matches 173. .261 of consensus"
11758..12719
/notes="CpG island"
/evidence="not_experimental"
12259..12372
/notes="57 copies 2 mer cc 66% conserved"
12810..13093
/notes="AluSg repeat: matches 3. .290 of consensus"
13220..13427
/notes="L2 repeat: matches 2414. .2615 of consensus"
13434..13624
/notes="MER53 repeat: matches 1. .189 of consensus"
13693..13998
/notes="AluSx repeat: matches 1. .306 of consensus"
14150..14201
/notes="MIR repeat: matches 68. .125 of consensus"
14376..14517
/notes="MIR repeat: matches 100. .244 of consensus"
15679..15715


```

repeat_region /note="MIR repeat: matches 218. .254 of consensus"
16039. .16334
repeat_region /note="AluJo repeat: matches 18. .311 of consensus"
16409. .16636
repeat_region /note="MIR repeat: matches 32. .261 of consensus"
16832. .17009
repeat_region /note="MIR repeat: matches 51. .244 of consensus"
17818. .18085
repeat_region /note="L2 repeat: matches 2311. .2605 of consensus"
18626. .18913
repeat_region /note="AluSx repeat: matches 1. .291 of consensus"
18925. .19128
repeat_region /note="L2 repeat: matches 2324. .2555 of consensus"
19202. .19370
repeat_region /note="AluSg/x repeat: matches 134. .302 of consensus"
19372. .19665
repeat_region /note="AluSg repeat: matches 1. .293 of consensus"
19679. .19986
repeat_region /note="AluSg repeat: matches 2. .308 of consensus"
complement(20236. .20676)
misc_feature /genes="dJ20208.1"
/note="match: GSS: Em:B43196"
complement(20501. .20541)
misc_feature /genes="dJ20208.1"
/note="match: GSS: Em:B43451"
20905. .21014
repeat_region /note="L2 repeat: matches 2593. .2706 of consensus"
21016. .21075
repeat_region /note="MIR repeat: matches 130. .191 of consensus"
21171. .21501
repeat_region /note="AluSx repeat: matches 1. .311 of consensus"
22030. .22214

Query Match 6.4%; Score 63; DB 9; Length 143065;
Best Local Similarity 100.0%; Pred. No. 1.7e-23;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 39 TCTGACCTCGTATCGCCGCCCTCAGCTTCCAAAGTCTCGGATTACAGCATGAC 98
Db 19592 TCTGACCTCGTATCGCCGCCCTCAGCTTCCAAAGTCTCGGATTACAGCATGAC 19651

QY 99 CAC 101
Db 19652 CAC 19654

HS57G9 113872 bp DNA linear PRI 05-JUN-2003
Human DNA sequence from clone CTA-57G9 on chromosome 22q12.1,
complete sequence.
ACCESSION Z95116
VERSION Z95116.1 GI:2832592
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 113872)
McLaren.S.
Direct Submission
Submitted (05-JUN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 5, 1998 this sequence version replaced gi:2578126.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:

```

```

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep CTA-57G9 is from
the human BAC library described in U-J. Kim et al. (1996) Genomics
34, 213-218.
VECTOR: pBAC108L
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----

```

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e. phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

This sequence was generated from part of bacterial clone contigs of human chromosome 22, constructed by the Sanger Centre Chromosome 22 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr22>

This sequence is the entire insert of clone CTA-57G9 The true right end of clone RP5-117M9 is at 100 in this sequence.

FEATURES

```

Location/Qualifiers
1. 113872
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="22"
/map="q12.1"
/clone="CTA-57G9"
/clone_lib="CIT978SK-A1"
17. 392
/note="match: GSS: Em:B14225"
repeat_region /note="2.0 copies 6 mer GCCATG 24% conserved"
32. 43
repeat_region /note="3.3 copies 3 mer ATC 20% conserved"
171. 180
complement(305. 682)
/note="match: GSS: Em:B14096"
misc_feature complement(387. 678)
/note="match: GSS: Em:B14098"
misc_feature complement(430. 682)
/note="match: GSS: Em:B36168"
593. 618
repeat_region /note="13.0 copies 2 mer GT 43% conserved"
740. 1035
repeat_region /note="Alu repeat: matches 1. .296 of consensus"
1177. 1188
repeat_region /note="2.0 copies 6 mer GAAAGG 24% conserved"
complement(1326. 1866)
/note="MLT1F2 repeat: matches 33. .554 of consensus"
complement(1867. 2202)
repeat_region /note="MLT1A repeat: matches 24. .374 of consensus"
complement(2221. 2519)
repeat_region /note="MER21C repeat: matches 666. .932 of consensus"
2520. 2659
repeat_region /note="FLAM C repeat: matches 1. .129 of consensus"
complement(2660. 3073)
repeat_region /note="MER21C repeat: matches 21. .666 of consensus"
3241. 3452
repeat_region /note="MLT1D repeat: matches 1. .236 of consensus"
3461. 4005
/note="13.6 copies 40 mer
AGAGTGATGATGAAGGAAACAGGAGGAGAGGTGATGA 780% conserved"
4013. 4345
repeat_region /note="MLT1D repeat: matches 180. .505 of consensus"
4553. 4566
repeat_region /note="14.0 copies 1 mer T 28% conserved"
4615. 4629
repeat_region

```

```

/note="3.0 copies 5 mer CAAA 21% conserved"
join(4811..4935,8717..8870,20796..21128,22078..22209)
/genes="BX5709.1"
CDS
join(4811..4935,8717..8870,20796..21128,22078..22209)
/genes="BX5709.1"
/note="supported by GENSCAN"
match: ESTs: Em:AA506741 Em:AA670682
match: proteins: Tr:P91972 Sw:P98065 Tr:Q15113 Tr:Q35113
Tr:Q24132 Tr:Q70244 Tr:O57433 Tr:O88204 Sw:P97435
Tr:O60494 Tr:O57658 Tr:O57074 Tr:O57434 Tr:O08628
Tr:Q61398 Sw:P98066 Tr:Q23995 Sw:P97333 Tr:O08859
Tr:O95917"
/codon_start=3
/evidence=not experimental
/product="BX5709.1 (novel Kringle and CUB domain protein)"
/protein_id="CAB62952.1"
/db_xref="GI:6572252"
/translation="PGLGICYDHGNPPLTGTSTKSLTQTICISFCRSQRFKAG
MESGYACFCGNPDYKYGKGAASCTCNVCFGDHTQPCGGDRIILFTLVGACGNY
SAMSSVYSPDPDVTATGVCVWTIRVPGASHIHFSPLFDIRDSADWVLLDGYTH
EVLARFHGSRPLSNVSLDVLVLYPFSDRINQAGFAVLQAVKEELPQRPVAVQ
TVAEVLTEQNLSVSAARSKVLVIITS"
4995..5311
/note="LIMB7 repeat: matches 5850..6194 of consensus"
5333..6111
/note="LIME1 repeat: matches 5043..5827 of consensus"
6112..6420
/note="ALUSg repeat: matches 1..306 of consensus"
6421..6501
/note="LIME1 repeat: matches 5827..5906 of consensus"
6502..6965
/note="LIME1 repeat: matches 4987..5467 of consensus"
6966..7272
/note="ALUSx repeat: matches 3..311 of consensus"
7273..7946
/note="LIME1 repeat: matches 5467..6068 of consensus"
complement(7947..8246)
/note="ALUSx repeat: matches 1..300 of consensus"
8247..8291
/note="LIME1 repeat: matches 6068..6113 of consensus"
8327..8426
/note="MIR repeat: matches 134..262 of consensus"
8941..8952
/note="3.0 copies 4 mer CCT 24% conserved"
8960..8972
/note="3.2 copies 4 mer ATTC 26% conserved"
complement(8997..9100)
/note="L2 repeat: matches 3138..3259 of consensus"
9244..9253
/note="10.0 copies 1 mer T 20% conserved"
9570..9582
/note="13.0 copies 1 mer A 26% conserved"
complement(9708..10119)
/note="MLT1C repeat: matches 8..465 of consensus"
complement(10556..10724)
/note="MIR repeat: matches 74..252 of consensus"
10941..10916
/note="LIMB7 repeat: matches 5508..5593 of consensus"
10917..11122
/note="ALUSg/x repeat: matches 83..296 of consensus"
11123..11661
/note="LIMB7 repeat: matches 5593..6122 of consensus"
11710..11811
/note="Tiggers repeat: matches 2307..2406 of consensus"
11816..11834
/note="2.1 copies 9 mer TTAATAATG 38% conserved"
complement(12111..12286)
/note="MIR repeat: matches 47..262 of consensus"
12663..12810
/note="ALUSg/x repeat: matches 139..230 of consensus"
complement(13522..13690)
/note="MIR repeat: matches 78..252 of consensus"
complement(13684..13732)

```

```

/note="L2 repeat: matches 3182..3229 of consensus"
complement(13805..13964)
/note="MER104 repeat: matches 1..176 of consensus"
14449..14464
/note="2.3 copies 7 mer GTTAAAG 23% conserved"
complement(14535..14785)
/note="MIR repeat: matches 5..259 of consensus"
complement(14788..14982)
/note="LIME1 repeat: matches 6741..6952 of consensus"
14983..15002
/note="10.0 copies 2 mer TA 40% conserved"
15003..15057
/note="LTR39-int repeat: matches 3905..3959 of consensus"
15067..15093
/note="13.5 copies 2 mer TA 36% conserved"
15069..15091
/note="2.9 copies 8 mer TATATATG 46% conserved"
15100..15189
/note="LIMC4 repeat: matches 6614..6703 of consensus"
15158..15230
/note="LIMCa repeat: matches 1995..2064 of consensus"
complement(15233..15319)
/note="LIME1 repeat: matches 6659..6744 of consensus"
complement(15320..15625)
/note="ALUSp repeat: matches 1..307 of consensus"
complement(15626..16420)
/note="LIME1 repeat: matches 5869..6659 of consensus"
16421..16729
/note="ALUy repeat: matches 1..309 of consensus"

Query Match 6.1%; Score 60; DB 9; Length 113872;
Best Local Similarity 100.0%; Pred. No. 7.8e-22;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 GTGATCCGCGCGCTCCCAAGTCTGGGATTACAGCATGAGCCACCGCGCCC 108
|||||
DB 112880 GTGATCCGCGCGCTCCCAAGTCTGGGATTACAGCATGAGCCACCGCGCCC 112821

RESULT 10
AC004645/c
LOCUS AC004645 44174 bp DNA linear PRI 01-MAY-1998
DEFINITION Homo sapiens chromosome 16, cosmid clone 400C4 (LANL), complete
sequence.
ACCESSION AC004645
VERSION AC004645.1 GI:3097839
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (Bases 1 to 44174)
Ricke,D.O.
Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
Unpublished
2 (Bases 1 to 44174)
Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
Misra,M. and Deaven,L.
Sequencing of Human Chromosome 16p13.3
Unpublished
3 (Bases 1 to 44174)
Ricke,D.O., Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E.,
Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S.,
Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,J.,
White,S., Ueng,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M.,
Misra,M. and Deaven,L.
Direct Submission
Submitted (01-MAY-1998) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los

```

```

FEATURES             Alamos, NM 87545, USA
    source            Location/Qualifiers
1..44174
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="16"
    /map="16p13.3"
    /clone="400C4"
    complement(355..1214)
    /standard_name="fmc17"
    /note="96% identity dbEST:AA631950"
416..458
    /note="GRAIL 2 excellent exon, frame 2"
962..1020
    /note="GRAIL 2 excellent exon, frame 0"
1396..1508
    /note="87% identity no80a12.s1"
    /db_xref="dbEST:AA604650"
    complement(1867..2144)
    /note="87% identity X59618"
    complement(3192..3576)
    /rpt_family="Alu"
    complement(3618..4137)
    /rpt_family="Alu"
4278..4599
    /rpt_family="Alu"
4589..4588
    /note="(A)20"
    /rpt_type=tandem
    /rpt_unit="A"
5384..5837
    /note="99% identity EST z131a04.s1"
    /db_xref="dbEST:AA706919"
    complement(6884..7023)
    /rpt_family="Alu"
    complement(7093..7725)
    /rpt_family="Alu"
7811..8117
    /rpt_family="Alu"
9238..9330
    /note="GRAIL 2 excellent exon, frame 1"
    complement(10614..10716)
    /note="GRAIL 2 excellent exon, frame 2"
    complement(11928..12071)
    /rpt_family="Alu"
13214..13542
    /rpt_family="Alu"
13711..13750
    /note="(AC)20"
    /rpt_type=tandem
    /rpt_unit="AC"
14070..14157
    /rpt_family="L1MCB"
14467..14709
    /rpt_family="L1M4"
    complement(14704..15013)
    /rpt_family="Alu"
    complement(15018..15315)
    /rpt_family="Alu"
15367..15604
    /rpt_family="L1MCB"
    complement(16360..16708)
    /rpt_family="Alu"
16375..16397
    /note="(T)23"
    /rpt_type=tandem
    /rpt_unit="T"
16724..16794
    /rpt_family="Alu"
16784..16996
    /rpt_family="7SL"
16978..17327
    /rpt_family="Alu"

repeat_region       /rpt_family="Alu"
17328..17635
    /rpt_family="Alu"
17745..18053
    /rpt_family="Alu"
    complement(18227..18533)
    /rpt_family="Alu"
18809..18917
    /rpt_family="L1M2"
18954..19282
    /rpt_family="Alu"
19278..19798
    /rpt_family="L1ME"
19948..20656
    /rpt_family="Alu"
20650..20833
    /rpt_family="Alu"
    complement(20981..21127)
    /rpt_family="Alu"
21330..21684
    /rpt_family="Alu"
21714..22048
    /rpt_family="Alu"
22060..22632
    /rpt_family="L1"
    complement(22661..22976)
    /rpt_family="Alu"
22661..22683
    /note="(T)23"
    /rpt_type=tandem
    /rpt_unit="T"
22994..23307
    /rpt_family="Alu"
23540..24242
    /rpt_family="L1"
23894..23916
    /note="(T)23"
    /rpt_type=tandem
    /rpt_unit="T"
24256..24578
    /rpt_family="Alu"
24635..24987
    /rpt_family="Alu"
    complement(25737..26036)
    /rpt_family="Alu"
    complement(26209..26434)
    /rpt_family="Alu"
    complement(26443..26829)
    /rpt_family="Alu"
26859..27193
    /rpt_family="Alu"
    complement(27569..27877)
    /rpt_family="Alu"
    complement(27852..28119)
    /rpt_family="L1"
    complement(28107..28242)
    /rpt_family="Alu"
    complement(28199..28953)
    /rpt_family="L1"
29073..29237
    /rpt_family="Alu"
    complement(29616..29932)
    /rpt_family="Alu"
30472..31073
    /rpt_family="Alu"
    complement(31309..31412)
    /rpt_family="Alu"
31636..31943
    /rpt_family="Alu"
32031..32566
    /rpt_family="Alu"
32519..33006
    /rpt_family="Alu"

```

```

repeat_region complement(33104..33431)
/rpt_family="Alu"
repeat_region complement(33411..33770)
/rpt_family="L1"
repeat_region complement(34043..34190)
/rpt_family="Alu"
repeat_region complement(34498..34794)
/rpt_family="Alu"
repeat_region 34501..34527
/notes="(T)27"
/rpt_type=tandem
/rpt_unit="T"
repeat_region 34797..35210
/rpt_family="L1"
repeat_region complement(35186..35517)
/rpt_family="Alu"
repeat_region 37067..37493
/rpt_family="MER74B"
repeat_region 37452..38121
/rpt_family="Alu"
repeat_region 37923..37945
/notes="(A)23"

Query Match 6.0%; Score 59; DB 9; Length 44174;
Best Local Similarity 100.0%; Pred. No. 2.7e-21;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 CGATCTCTGACCTGCTGATCCCGCGCTCAGCTTCCCAAGTGTGGGATTACAGGC 92
Db 23075 CGATCTCTGACCTGCTGATCCCGCGCTCAGCTTCCCAAGTGTGGGATTACAGGC 23017

RESULT 11
AL157412/c
LOCUS AL157412 45454 bp DNA linear PRI 31-JUL-2000
DEFINITION Human DNA sequence from clone RP11-125P18 on chromosome 20 Contains
STSS and GSSs, complete sequence.
ACCESSION AL157412
VERSION AL157412.16 GI:8894235
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 45454)
Wilson,S.
REFERENCE Direct Submission
AUTHORS Submitted (27-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
TITLE CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
JOURNAL requests: clonerequests@sanger.ac.uk
COMMENT On Jul 1, 2000 this sequence version replaced gi:8546617.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
En: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
RP11-125P18 is from the library RPC1-11.1 constructed at the

```

Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/> VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-125P18 It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP5-1097P24 is at 45355 in this sequence. The true right end of clone RP5-1185K9 is at 100 in this sequence.

```

Location/Qualifiers
FEATURES
    source
        1..45454
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="20"
            /clone="RP11-125P18"
            /clone_lib="RPC1-11.1"
            /complement(630..1120)
            /note="match: GSS: Em:AQ155865"
            complement(638..1131)
            /note="match: GSS: Em:AQ727438"
            831..966
            /note="MIR repeat: matches 49..182 of consensus"
            966..11013
            /note="3 copies 16 mer 93% conserved"
            1139..1633
            /note="match: GSS: Em:AQ484484"
            2474..2555
            /note="41 copies 2 mer tt 63% conserved"
            3028..3285
            /note="MLT1J repeat: matches 244..508 of consensus"
            complement(3139..3392)
            /note="match: GSS: Em:AQ388774"
            complement(3200..3582)
            /note="match: GSS: Em:AQ972678"
            3318..3393
            /note="38 copies 2 mer ta 67% conserved"
            complement(3320..3573)
            /note="match: GSS: Em:AL219668"
            3327..3554
            /note="match: STS: Em:AQ024971"
            complement(3337..3578)
            /note="match: GSS: Em:AL241313"
            complement(3340..3582)
            /note="match: GSS: Em:AQ348189"
            3345..3543
            /note="match: STS: Em:G09236"
            complement(3346..3572)
            /note="match: STS: Em:AQ029014"
            3351..3580
            /note="match: GSS: Em:AL229216"
            3352..3522
            /note="match: GSS: Em:AL285765"
            3360..3534
            /note="match: GSS: Em:AL193675"
            3362..3473
            /note="7 copies 16 mer 77% conserved"
            3364..3534
            /note="match: GSS: Em:AQ527877"
            3371..3474
            /note="26 copies 4 mer atgg 77% conserved"
            complement(3375..3539)
            /note="match: GSS: Em:AL185783"
            3397..3583
            /note="match: GSS: Em:AL010814"
            complement(3402..3545)
            /note="match: STS: Em:G08408"
            complement(3428..3580)
            /note="match: GSS: Em:AL236234"
            complement(3452..3586)
            /note="match: STS: Em:Li8007"
            3452..3580
            /note="match: GSS: Em:AL273514"

```

misc_feature 3454. .3585 /note="match: GSS: Em:AL276849"
repeat_region 3490. .3581 /note="23 copies 4 mer tggg 84% conserved"
repeat_region 3581. .3532 /note="THE1A repeat: matches 1. .353 of consensus"
misc_feature complement(4751. .5369)
misc_feature /note="match: GSS: Em:B53127"
repeat_region complement(5089. .5368)
repeat_region /note="match: GSS: Em:AO242433"
repeat_region 5509. .5689 /note="MER5A repeat: matches 1. .186 of consensus"
repeat_region 6922. .7098 /note="MIR repeat: matches 46. .228 of consensus"
misc_feature 7262. .7369 /note="match: STS: Em:HSPF1266"
repeat_region 7636. .7590 /note="LTR33 repeat: matches 172. .514 of consensus"
misc_feature 7980. .8562 /note="match: GSS: Em:AO535194"
misc_feature 7996. .8542 /note="match: GSS: Em:AO548736"
repeat_region 8426. .8473 /note="12 copies 4 mer gaag 85% conserved"
repeat_region 9632. .9679 /note="12 copies 4 mer tata 81% conserved"
repeat_region 9678. .9719 /note="21 copies 2 mer at 83% conserved"
repeat_region 9684. .9719 /note="9 copies 4 mer atat 88% conserved"
repeat_region 11764. .11946 /note="MER5A repeat: matches 1. .189 of consensus"
repeat_region 13779. .13845 /note="MER5A repeat: matches 5. .71 of consensus"
repeat_region 13917. .14215 /note="AluX repeat: matches 1. .299 of consensus"
misc_feature 16015. .16486 /note="match: GSS: Em:AQ147161"
repeat_region 16019. .16319 /note="AluX repeat: matches 1. .307 of consensus"
repeat_region 16442. .16595 /note="MER5A repeat: matches 9. .189 of consensus"
repeat_region 18358. .18757 /note="LMC3 repeat: matches 7324. .7733 of consensus"
misc_feature 18729. .19104 /note="match: GSS: Em:AQ184417"
repeat_region 18846. .19024 /note="MIR repeat: matches 31. .197 of consensus"
repeat_region 19423. .19519 /note="MER5A repeat: matches 64. .168 of consensus"
repeat_region 19668. .19876 /note="MER58A repeat: matches 1. .224 of consensus"
repeat_region 19938. .20058 /note="MER5A repeat: matches 55. .188 of consensus"
repeat_region 20212. .20608 /note="MIR repeat: matches 3. .390 of consensus"
misc_feature 20523. .21105 /note="match: GSS: Em:AQ390380"
repeat_region 20792. .21253 /note="LIME1 repeat: matches 5708. .6158 of consensus"
repeat_region 21283. .22286 /note="LIME1 repeat: matches 4597. .5629 of consensus"
repeat_region 22523. .24133 /note="LIME5 repeat: matches 4497. .6175 of consensus"
repeat_region 24144. .25028 /note="LIPAG repeat: matches 5259. .6142 of consensus"
repeat_region 25028. .25322 /note="LIPAG repeat: matches 4945. .5239 of consensus"
repeat_region 25357. .26137 /note="LIP4 repeat: matches 2765. .3567 of consensus"
repeat_region 26138. .26440 /note="AluSg repeat: matches 1. .295 of consensus"
repeat_region 26441. .26763

misc_feature /note="LIM4 repeat: matches 2448. .2765 of consensus"
repeat_region complement(26705. .26933)
repeat_region /note="match: GSS: Em:AQ245856"
repeat_region 27193. .27364 /note="LIM4 repeat: matches 2309. .2474 of consensus"
repeat_region 27365. .27864 /note="LIP2 repeat: matches 5646. .6154 of consensus"
repeat_region 27865. .27900 /note="LIM4 repeat: matches 2273. .2309 of consensus"
repeat_region 27901. .28132 /note="58 copies 4 mer atat 59% conserved"
repeat_region 27937. .28080 /note="9 copies 16 mer f3% conserved"
misc_feature complement(28031. .28412)
misc_feature /note="match: GSS: Em:A2001959"

Query Match 6.0%; Score 59; DB 9; Length 45454;
Best Local Similarity 100.0%; Pred.No. 2.7e-21;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 CGATCTCTGACCTGCTGATCCGCCGCTCAGCTTCCCAAAGTCTGGGATTACAGGC 92
|||||
Db 43027 CGATCTCTGACCTGCTGATCCGCCGCTCAGCTTCCCAAAGTCTGGGATTACAGGC 42969

RESULT 12
LOCUS AF212832/c 122029 bp DNA linear PRI 19-NOV-2002
DEFINITION Homo sapiens chromosome 8 clone CTC-806C5 map 8q24.3, complete
sequence.
ACCESSION AF212832
VERSION AF212832.5 GI:25100909
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (Bases 1 to 122029)
AUTHORS Blechschmidt,K., Schattevoy,R., Baumgart,C. and Rosenthal,A.
JOURNAL Unpublished
REFERENCE 2 (Bases 1 to 122029)
AUTHORS Blechschmidt,K., Wen,G., Schilhabel,M., Baumgart,C., Menzel,U.,
Dette,M. and Rosenthal,A.
JOURNAL Direct Submission
REFERENCE 3 (Bases 1 to 122029)
AUTHORS Lagemann,D. and Platzter,M.
JOURNAL Direct Submission
REFERENCE 4 (Bases 1 to 122029)
AUTHORS Lagemann,D. and Platzter,M.
JOURNAL Submitted (10-SEP-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
REFERENCE 5 (Bases 1 to 122029)
AUTHORS Lagemann,D. and Platzter,M.
JOURNAL Submitted (19-NOV-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
COMMENT On Nov 19, 2002 this sequence version replaced gi:22773267.
----- Genome Center
Center: Insitute of Molecular Biotechnology
Center code: IMB
Web site: http://genome.imb-jena.de/
Contact: gscj-submit@genome.imb-jena.de
----- Project Information
Center project name: H283
Center clone name: CTC-806C5
----- Summary Statistics
Sequencing vector: pUC18; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 121876 bases at least Q40
Consensus quality: 121937 bases at least Q30
Consensus quality: 121979 bases at least Q20
Quality coverage: 10.81x

REFERENCE

AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 TITLE 1 (bases 1 to 123291)
 JOURNAL Sulston, J.E. and Waterston, R.
 MEDLINE Toward a complete human genome sequence
 PUBMED Genome Res. 8 (11), 1097-1108 (1998)
 99063792

REFERENCE

AUTHORS 2 (bases 1 to 123291)
 TITLE Desai, A., Kozlowski, A. and Boyer, E.
 JOURNAL The sequence of Homo sapiens BAC clone RP11-279K24
 MEDLINE Unpublished (2001)
 PUBMED 3 (bases 1 to 123291)
 9847074

REFERENCE

AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (03-DEC-2001) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE

AUTHORS 4 (bases 1 to 123291)
 TITLE Waterston, R.H.
 JOURNAL Direct Submission
 MEDLINE Submitted (24-APR-2002) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

REFERENCE

AUTHORS 5 (bases 1 to 123291)
 TITLE Waterston, R.
 JOURNAL Direct Submission
 MEDLINE Submitted (29-MAY-2002) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 On Apr 24, 2002 this sequence version replaced gi:18030153.

COMMENT

----- Genome Center
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: <http://genome.wustl.edu/gsc>
 Contact: sapiens@watson.wustl.edu
 ----- Summary Statistics
 Center project name: H_NH0279K24
 Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frenken, S., Tateno, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
 VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-45120, 2000 bp overlap; the clone sequenced to the right is RP11-173M11. Actual start of this clone is at base position 108871 of RP11-45120; actual end is at base position 123291 of RP11-279K24.

Unresolved tandem repeats exist between 44681 and 46316.
 Polymorphisms exist between AC096659, AC0110771 and AC104070. Data from AC110771 was used to finish AC104070.

The sequence of AC068461 has been incorporated into AC104070.

FEATURES

Source	Location/Qualifiers
1. .123291	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="4"
	/map="4"
	/clone="RP11-279K24"
	/clone_lib="RPC1-11"
repeat_region	174..350
	/rpt_family="MER1_type"
repeat_region	1085..1129
	/rpt_family="(CA)n"
repeat_region	2338..3252
	/rpt_family="Alu"
repeat_region	3691..3908
	/rpt_family="L2"
repeat_region	4040..4334
	/rpt_family="Alu"
repeat_region	4345..4522
	/rpt_family="Alu"
repeat_region	4525..5166
	/rpt_family="ERV1"
repeat_region	5269..5343
	/rpt_family="MIR"
repeat_region	5469..5535
	/rpt_family="MIR"
repeat_region	5866..6417
	/rpt_family="ERV1"
repeat_region	7027..7321
	/rpt_family="L2"
repeat_region	7331..7424
	/rpt_family="L2"
repeat_region	7535..7724
	/rpt_family="ERV1"
repeat_region	7739..8079
	/rpt_family="ERV1"
repeat_region	11056..11367
	/rpt_family="Alu"
repeat_region	12482..12789
	/rpt_family="Alu"
repeat_region	12790..12814
	/rpt_family="AT-rich"
repeat_region	13276..13375
	/rpt_family="L2"
repeat_region	13504..13774
	/rpt_family="L1"
repeat_region	13776..14086
	/rpt_family="Alu"
repeat_region	14090..14277
	/rpt_family="L1"
repeat_region	14307..14427
	/rpt_family="L1"
repeat_region	14642..14709
	/rpt_family="U4"
repeat_region	14802..14883
	/rpt_family="CT-rich"
repeat_region	15149..15220
	/rpt_family="ERV1"
repeat_region	15221..15419
	/rpt_family="MaLR"
repeat_region	15420..16124
	/rpt_family="ERV1"
repeat_region	16124..16356
	/rpt_family="MaLR"
repeat_region	16357..16729
	/rpt_family="ERV1"

```

repeat_region 16730..17029
/rpt_family="Alu"
repeat_region 17030..17494
/rpt_family="ERV1"
repeat_region 17517..17607
/rpt_family="L2"
repeat_region 18972..19151
/rpt_family="ERV1"
repeat_region 19204..19453
/rpt_family="ERV1"
repeat_region 19564..19640
/rpt_family="GA-rich"
repeat_region 20193..21044
/rpt_family="ERV1"
repeat_region 21501..21792
/rpt_family="Alu"
repeat_region 23724..23747
/rpt_family="AT-rich"
repeat_region 24122..24419
/rpt_family="Alu"
repeat_region 24525..24920
/rpt_family="MaLR"
repeat_region 25364..25788
/rpt_family="ERV1"
repeat_region 26725..27020
/rpt_family="Alu"
repeat_region 27333..27551
/rpt_family="MIR"
repeat_region 28446..28951
/rpt_family="Alu"
repeat_region 28952..29088
/rpt_family="Alu"
repeat_region 29204..29740
/rpt_family="L1"
repeat_region 29741..30050
/rpt_family="Alu"
repeat_region 30051..30244
/rpt_family="L1"
repeat_region 30245..30462
/rpt_family="Alu"
repeat_region 30463..30556
/rpt_family="L1"
repeat_region 30733..30767
/rpt_family="U2"
repeat_region 30768..30845
/rpt_family="L1"

Query Match 6.0%; Score 59; DB 9; Length 123291;
Best Local Similarity 100.0%; Pred. No. 2.8e-21;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 34 CGATCTCCTGACCTCGATCGCGCCCTCAGCTTCCCAAGTGTGGGATTACAGGC 92
|||||
Db 43370 CGATCTCCTGACCTCGATCGCGCCCTCAGCTTCCCAAGTGTGGGATTACAGGC 43428

RESULT 14
BX571818/c
LOCUS
DEFINITION Homo sapiens chromosome 1 clone RP11-39J22, WORKING DRAFT SEQUENCE,
5 unordered pieces.
ACCESSION BX571818
VERSION BX571818.2 GI:33386608
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 148963)
McLay, K.
Direct Submission
TITLE Submitted (30-JUL-2003) Wellcome Trust Sanger Institute, Hinxton,
JOURNAL Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

```

COMMENT

humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Jul 31, 2003 this sequence version replaced gi:33086351.
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk
 ----- Project Information
 Center project name: BA38J22
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Chemistry: Dye-terminator; 100% of reads
 Consensus quality: 147613 bases at least Q40
 Consensus quality: 147851 bases at least Q30
 Consensus quality: 148031 bases at least Q20
 Insert size: 148563; sum-of-contigs
 Insert size: 148899; 19.7% error; agarose-fp
 Quality coverage: 24.62x in Q20 bases; sum-of-contigs Quality
 coverage: 25.09x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

* 1 2276: contig of 2276 bp in length
 * 2277 2376: gap of 100 bp
 * 2377 35686: contig of 33310 bp in length
 * 35687 35786: gap of 100 bp
 * 35787 51485: contig of 15699 bp in length
 * 51486 51585: gap of 100 bp
 * 51586 132546: contig of 80961 bp in length
 * 132547 132646: gap of 100 bp
 * 132647 148963: contig of 16317 bp in length.

FEATURES

source

1..148963
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-38J22"
 /clone_lib="RPC1-11.1"
 1..2276
 /note="assembly fragment:00367"
 fragment_chain:1
 clone_end:SP6
 vector_side:left
 2377..35686
 /note="assembly fragment:04519"
 fragment_chain:1
 35787..51485
 /note="assembly fragment:03291"
 fragment_chain:1
 51586..132546
 /note="assembly fragment:03347"
 fragment_chain:1
 132547..148963
 /note="assembly fragment:04446"
 fragment_chain:1
 clone_end:T7
 vector_side:right

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

ORIGIN

Query Match 6.0%; Score 59; DB 2; Length 148963;
 Best Local Similarity 100.0%; Pred. No. 2.9e-21;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 GTGATCCCGCCGCTCAGCTTCCCAAGTGTGGGATTACAGGCACCGCC 107
 |||||
 Db 15354 GTGATCCCGCCGCTCAGCTTCCCAAGTGTGGGATTACAGGCACCGCC 15296

RESULT 15

AC012440

LOCUS

DEFINITION

AC012440 Homo sapiens clone RP11-11H11, WORKING DRAFT SEQUENCE, 21 unordered

pieces.

ACCESSION

AC012440

VERSION

AC012440.2 GI:7137115

KEYWORDS

HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1 (bases 1 to 159619)

AUTHORS

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

TITLE

Homo sapiens, clone RP11-11H11

REFERENCE

Unpublished

AUTHORS

2 (bases 1 to 159619)

REFERENCE

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,

Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Bouckghalter,B.,

Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,

Cooke,P., DeArelano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,

Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D.,

Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,

McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J.,

Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,

Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,

Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,

Tejeda-S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,

Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.

TITLE

Direct Submission

JOURNAL

Submitted (27-OCT-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Mar 1, 2000 this sequence version replaced gi:6136376.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research-

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L3166

Center Clone name: 11 H.11

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 149892 bases at least Q40

Consensus quality: 154478 bases at least Q30

Consensus quality: 156029 bases at least Q20

Insert size: 165000; agarose-gel

Insert size: 157619; sum-of-contigs

Quality coverage: 4.7 in Q20 bases; agarose-gel

Quality coverage: 4.9 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently

* consists of 21 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 1125: contig of 1125 bp in length

* 1126 1225: gap of 100 bp

* 1226 2918: contig of 1693 bp in length

* 2919 3019: gap of 100 bp

* 3019 4446: contig of 1428 bp in length

*

*	4447	4546: gap of 100 bp
*	4547	5800: contig of 1254 bp in length
*	5801	5900: gap of 100 bp
*	5901	6771: contig of 871 bp in length
*	6772	6871: gap of 100 bp
*	6872	8592: contig of 1721 bp in length
*	8593	8692: gap of 100 bp
*	8693	11437: contig of 2745 bp in length
*	11438	11537: gap of 100 bp
*	11538	14484: contig of 2947 bp in length
*	14485	14584: gap of 100 bp
*	14585	17422: contig of 2838 bp in length
*	17423	17522: gap of 100 bp
*	17523	21714: contig of 4192 bp in length
*	21715	21814: gap of 100 bp
*	21815	27645: contig of 5832 bp in length
*	27647	27745: gap of 100 bp
*	27747	33347: contig of 5601 bp in length
*	33348	33447: gap of 100 bp
*	33448	41489: contig of 8042 bp in length
*	41490	41589: gap of 100 bp
*	41590	49987: contig of 8398 bp in length
*	49988	50087: gap of 100 bp
*	50088	57308: contig of 7219 bp in length
*	57309	57406: gap of 100 bp
*	57407	63803: contig of 6397 bp in length
*	63804	63903: gap of 100 bp
*	63904	74106: contig of 10203 bp in length
*	74107	74206: gap of 100 bp
*	74207	85991: contig of 11785 bp in length
*	85992	86091: gap of 100 bp
*	86092	104012: contig of 17921 bp in length
*	104013	104112: gap of 100 bp
*	104113	124486: contig of 20374 bp in length
*	124487	124586: gap of 100 bp
*	124587	159619: contig of 35033 bp in length.

FEATURES

source

1..159619	Location/Qualifiers
/organism="Homo sapiens"	
/mol_type="genomic DNA"	
/db_xref="taxon:9606"	
/clone="RP11-11H11"	
/clone_lib="RPCI-11 Human Male BAC"	
1..1125	
/note="assembly_fragment"	
1226..2918	
/note="assembly_fragment"	
3019..4446	
/note="assembly_fragment"	
4547..5800	
/note="assembly_fragment"	
5901..6771	
/note="assembly_fragment"	
clone_end:T7	
vector_side:right	
6872..8592	
/note="assembly_fragment"	
8693..11437	
/note="assembly_fragment"	
11538..14484	
/note="assembly_fragment"	
14585..17422	
/note="assembly_fragment"	
17523..21714	
/note="assembly_fragment"	
21815..27646	
/note="assembly_fragment"	
27747..33347	
/note="assembly_fragment"	
33448..41489	
/note="assembly_fragment"	
41590..49987	
/note="assembly_fragment"	

```
misc_feature      50088..57306
                  /note="assembly_fragment"
misc_feature      57407..63803
                  /note="assembly_fragment"
misc_feature      63904..74106
                  /note="assembly_fragment"
misc_feature      74207..85991
                  /note="assembly_fragment"
misc_feature      86092..104012
                  /note="assembly_fragment"
misc_feature      104113..124486
                  /note="assembly_fragment
                  clone_end:SP6
                  vector_side:left"
misc_feature      124587..159619
                  /note="assembly_fragment"

ORIGIN
Query Match      6.0%; Score 59; DB 2; Length 159619;
Best Local Similarity 100.0%; Pred. No. 2.9e-21;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      49  GTGATCGCGCCGCTCAGCTTCCCAAAGTGGTGGGATTACAGGCATGAGCCACCGCGCC 107
Db      75510 GTGATCGCGCCGCTCAGCTTCCCAAAGTGGTGGGATTACAGGCATGAGCCACCGCGCC 75568
```

Search completed: July 10, 2004, 14:37:04
Job time : 4191 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 10, 2004, 13:09:30 ; Search time 99 seconds
(without alignments)
5549.509 Million cell updates/sec

Title: US-09-972-032-1
Perfect score: 990
Sequence: 1 ggaatgtctcaggccaa.....aaatgagcgccgaagt 990

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0
Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfilesi.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	4.8	118067	4 US-09-497-855A-32	Sequence 32, Appl
2	45	4.5	505	4 US-09-621-976-15373	Sequence 15373, A
3	45	4.5	3267	2 US-08-257-963B-12	Sequence 12, Appl
4	45	4.5	3267	4 US-08-367-841A-12	Sequence 12, Appl
5	45	4.5	3267	5 PCT-US95-07201-12	Sequence 5, Appl
6	45	4.5	5262	4 US-08-520-373D-5	Sequence 43, Appl
7	45	4.5	22481	4 US-08-367-841A-43	Sequence 43, Appl
8	45	4.5	22481	5 PCT-US95-07201-43	Sequence 2, Appl
9	45	4.5	22484	4 US-09-875-223-2	Sequence 2, Appl
10	45	4.5	22484	4 US-09-875-114-2	Sequence 2, Appl
11	45	4.5	75395	4 US-09-984-890-3	Sequence 3, Appl
12	44	4.4	164	4 US-09-621-976-11863	Sequence 11863, A
13	44	4.4	281	4 US-09-621-976-1944	Sequence 1944, Ap
14	44	4.4	294	4 US-09-621-976-1939	Sequence 1939, Ap
15	44	4.4	294	4 US-09-621-976-1940	Sequence 1940, Ap
16	44	4.4	298	4 US-08-621-976-11526	Sequence 11526, A
17	44	4.4	298	4 US-08-621-976-13327	Sequence 13327, A
18	44	4.4	302	4 US-09-621-976-1937	Sequence 1937, Ap
19	44	4.4	302	4 US-09-621-976-1942	Sequence 1942, Ap
20	44	4.4	310	4 US-09-621-976-1943	Sequence 1943, Ap
21	44	4.4	312	4 US-09-621-976-12169	Sequence 12169, A
22	44	4.4	314	4 US-09-621-976-11935	Sequence 11935, A
23	44	4.4	314	4 US-09-621-976-11969	Sequence 11969, A
24	44	4.4	507	4 US-09-621-976-13632	Sequence 13632, A
25	44	4.4	99500	4 US-09-788-096-10	Sequence 10, Appl
26	44	4.4	392000	4 US-10-027-983-11	Sequence 11, Appl
27	43	4.3	4858	4 US-09-595-684B-28	Sequence 28, Appl

C 28 43 4.3 14364 4 US-10-067-443-20 Sequence 20, Appl
C 29 43 4.3 786431 4 US-09-751-389-3 Sequence 3, Appl
C 30 42 4.2 3350 3 US-09-110-116-2 Sequence 2, Appl
C 31 42 4.2 21784 4 US-09-820-002-3 Sequence 3, Appl
C 32 41 4.1 75395 4 US-09-984-890-3 Sequence 3, Appl
C 33 40 4.0 421 4 US-09-621-976-15130 Sequence 15130, A
C 34 40 4.0 438 4 US-09-621-976-15688 Sequence 15688, A
C 35 40 4.0 7210 2 US-08-257-963B-10 Sequence 10, Appl
C 36 40 4.0 7210 4 US-08-367-841A-10 Sequence 10, Appl
C 37 40 4.0 7210 5 PCT-US95-07201-10 Sequence 10, Appl
C 38 40 4.0 11811 3 US-09-078-294-7 Sequence 7, Appl
C 39 40 4.0 14581 4 US-08-520-373D-4 Sequence 4, Appl
C 40 40 4.0 20303 1 US-08-370-975B-6 Sequence 6, Appl
C 41 40 4.0 26764 1 US-08-370-975B-1 Sequence 1, Appl
C 42 40 4.0 70000 4 US-09-851-896-3 Sequence 3, Appl
C 43 40 4.0 112132 4 US-09-741-150-3 Sequence 3, Appl
C 44 40 4.0 112132 4 US-10-160-187-3 Sequence 3, Appl
C 45 40 4.0 152331 3 US-09-128-155-16 Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-09-497-855A-32
; Sequence 32, Application US/09497855A
; Patent No. 6605432
; GENERAL INFORMATION:
; APPLICANT: Huang, Tim
; TITLE OF INVENTION: HIGH-THROUGHPUT METHODS FOR DETECTING DNA METHYLATION
; FILE REFERENCE: UMO1523
; CURRENT APPLICATION NUMBER: US/09/497,855A
; CURRENT FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/120,592
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: 60/118,760
; PRIOR FILING DATE: 1999-02-05
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 32
; LENGTH: 118067
; TYPE: DNA
; ORGANISM: Homo sapiens;
US-09-497-855A-32

Query Match 4.8%; Score 48; DB 4; Length 118067;
Best Local Similarity 100.0%; Pred. No. 1.2e-11;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 49 GTGATCCGCCGCCCTCAGCTCCCAAGTGGTGGGATTACAGGCATGA 96
DB 33745 GTGATCCGCCGCCCTCAGCTCCCAAGTGGTGGGATTACAGGCATGA 33792

RESULT 2
US-09-621-976-15373
; Sequence 15373, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 15373
; LENGTH: 506
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-15373

Query Match 4.5%; Score 46; DB 4; Length 506;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 TCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 114
Db 76 TCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 121

RESULT 3

US-08-257-963B-12
; Sequence 12, Application US/08257963B
; Patent No. 5840686
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Becerra, S.
; APPLICANT: Patricia; Schwartz, Joan P.;
; APPLICANT: Taniwaki, Takayuki
; TITLE OF INVENTION: PIGMENT EPITHELIUM
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION OF ITS NOVEL
; TITLE OF INVENTION: BIOLOGICAL ACTIVITY AND SEQUENCES ENCODING
; TITLE OF INVENTION: AND EXPRESSING THE PROTEIN
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,963B
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/952,796
; FILING DATE: 24-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36434
; REFERENCE/DOCKET NUMBER: 20264126US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3267 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: JT109
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: 3.3 kb PCR product
; OTHER INFORMATION: using primers, SEQ ID No. 5840686 15 and 16

US-08-257-963B-12

Query Match 4.5%; Score 45; DB 2; Length 3267;
Best Local Similarity 100.0%; Pred. No. 2.8e-10;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 TCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 113
Db 625 TCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 669

RESULT 4

US-08-367-841A-12
; Sequence 12, Application US/08367841A
; Patent No. 6319687
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Rodriguez,
; APPLICANT: Ignacio R.; Mazuruk, Krzysztof;
; APPLICANT: Tombran-Tink, Joyce
; TITLE OF INVENTION: PIGMENT EPITHELIUM
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
; TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/367,841A
; FILING DATE: 30-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/257,963
; FILING DATE: 07-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/952,796
; FILING DATE: 24-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36434
; REFERENCE/DOCKET NUMBER: 20264126US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3267 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: JT109
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: 3.3 kb PCR product
; OTHER INFORMATION: using primers, SEQ ID No. 6319687 15 and 16

US-08-367-841A-12

Query Match 4.5%; Score 45; DB 4; Length 3267;
Best Local Similarity 100.0%; Pred. No. 2.8e-10;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 TCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 113
Db 625 TCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 669

RESULT 5

PCT-US95-07201-12
; Sequence 12, Application PC/TUS9507201
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Becerra, Sofia
; APPLICANT: Patricia; Schwartz, Joan P.;
; APPLICANT: Taniwaki, Takayuki
; TITLE OF INVENTION: PIGMENT EPITHELIUM

Qy 69 TCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 113
Db 625 TCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 669

RESULT 6

US-08-367-841A-12
; Sequence 12, Application US/08367841A
; Patent No. 6319687
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Rodriguez,
; APPLICANT: Ignacio R.; Mazuruk, Krzysztof;
; APPLICANT: Tombran-Tink, Joyce
; TITLE OF INVENTION: PIGMENT EPITHELIUM
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
; TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/367,841A
; FILING DATE: 30-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/257,963
; FILING DATE: 07-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/952,796
; FILING DATE: 24-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36434
; REFERENCE/DOCKET NUMBER: 20264126US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3267 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: Genomic DNA
; FEATURE:
; NAME/KEY: JT109
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: 3.3 kb PCR product
; OTHER INFORMATION: using primers, SEQ ID No. 6319687 15 and 16

US-08-367-841A-12

Query Match 4.5%; Score 45; DB 4; Length 3267;
Best Local Similarity 100.0%; Pred. No. 2.8e-10;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 69 TCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 113
Db 625 TCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 669

RESULT 7

PCT-US95-07201-12
; Sequence 12, Application PC/TUS9507201
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Becerra, Sofia
; APPLICANT: Patricia; Schwartz, Joan P.;
; APPLICANT: Taniwaki, Takayuki
; TITLE OF INVENTION: PIGMENT EPITHELIUM

Qy 69 TCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 113
Db 625 TCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCGGCC 669

TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morgan & Finnegan, L.L.P.
STREET: 345 Park Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07201
FILING DATE: 06-JUN-1995

CLASSIFICATION:
PRIORITY INFORMATION:
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

PRIORITY INFORMATION: 08/367,841
PRIORITY INFORMATION: 08/367,841

CURRENT FILING DATE: 1995-08-29
PRIOR APPLICATION NUMBER: 08/377,710
PRIOR FILING DATE: 1995-01-25
PRIOR APPLICATION NUMBER: 08/279,979
PRIOR FILING DATE: 1994-07-25
PRIOR APPLICATION NUMBER: 07/894,215
PRIOR FILING DATE: 1992-06-04
PRIOR APPLICATION NUMBER: 07/952,796
PRIOR FILING DATE: 1992-09-24
NUMBER OF SEQ ID NOS: 34
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 5

LENGTH: 5262
TYPE: DNA
ORGANISM: HUMAN

FEATURE:
OTHER INFORMATION: EXON 35-161; EXON 1142-1297; EXON 1984-2187;

OTHER INFORMATION: EXON 5170-5255; INTRON 162-1141; INTRON

OTHER INFORMATION: 1298-1983; INTRON 2188-5169; CDS 35-161; CDS

OTHER INFORMATION: 1142-1297; CDS 1984-2187; CDS 5170-5255

NAME/KEY: exon
LOCATION: (35)..(160)

NAME/KEY: exon
LOCATION: (1142)..(1297)

NAME/KEY: exon
LOCATION: (1984)..(2187)

NAME/KEY: exon
LOCATION: (5170)..(5256)

NAME/KEY: intron
LOCATION: (162)..(1141)

NAME/KEY: intron
LOCATION: (1298)..(1983)

NAME/KEY: intron
LOCATION: (2188)..(5169)

OTHER INFORMATION: n = a or g or t or c, any base

US-08-520-373D-5

Query Match 4.5%; Score 45; DB 4; Length 5262;

Best Local Similarity 100.0%; Pred. No. 2.7e-10;

Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCGCGGCC 113

DB 2619 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCGCGGCC 2663

RESULT 7

US-08-367-841A-43

Sequence 43, Application US/08367841A

Patent No. 6319587

GENERAL INFORMATION:

APPLICANT: Chader, Gerald J.; Rodriguez,

APPLICANT: Ignacio R.; Mazuruk, Krzysztof;

APPLICANT: Tombran-Tink, Joyce

TITLE OF INVENTION: PIGMENT EPITHELIUM

TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC

TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE

NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:

ADDRESSEE: Morgan & Finnegan

STREET: 345 Park Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy Disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/367,841A

FILING DATE: 30-DEC-1994

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/257,963
FILING DATE: 07-JUN-1994
PRIOR APPLICATION DATA: 07/952,796
FILING DATE: 24-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: DOROTHY R. AUTH
REGISTRATION NUMBER: 36434
REFERENCE/DOCKET NUMBER: 20264126US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 22481 Base Pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Unknown
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Pl-147
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: full length genomic
OTHER INFORMATION: sequence for PEDF plus flanking sequences.
US-08-367-841A-43

Query Match 4.5%; Score 45; DB 4; Length 22481;
Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 TCCCAAGTCTGGGATTACAGGATGAGCCACCGCGCCGCC 113
|||||
DB 17186 TCCCAAGTCTGGGATTACAGGATGAGCCACCGCGCCGCC 17230

RESULT 8
PCT-US95-07201-43
Sequence 43, Application PC/TUS9507201
GENERAL INFORMATION:
APPLICANT: Chader, Gerald J.; Becerra, Sofia
APPLICANT: Patricia; Schwartz, Joan P.;
APPLICANT: Taniwaki, Takayuki
TITLE OF INVENTION: PIGMENT EPITHELIUM
TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PEDF GENE
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morgan & Finnegan, L.L.P.
STREET: 345 Park Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07201
FILING DATE: 06-JUN-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/367,841
FILING DATE: 30-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/257,963
FILING DATE: 07-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/952,796

FILING DATE: 24-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: DOROTHY R. AUTH
REGISTRATION NUMBER: 36434
REFERENCE/DOCKET NUMBER: 20264126PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 22481 Base Pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Unknown
MOLECULE TYPE: Genomic DNA
FEATURE:
NAME/KEY: Pl-147
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: full length genomic
OTHER INFORMATION: sequence for PEDF plus flanking sequences.
PCT-US95-07201-43

Query Match 4.5%; Score 45; DB 5; Length 22481;
Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 TCCCAAGTCTGGGATTACAGGATGAGCCACCGCGCCGCC 113
|||||
DB 17186 TCCCAAGTCTGGGATTACAGGATGAGCCACCGCGCCGCC 17230

RESULT 9
US-09-875-223-2
Sequence 2, Application US/09875223
Patent No. 6391850
GENERAL INFORMATION:
APPLICANT: No. 6391850thwestern University
APPLICANT: No. 63918501 Bouck
APPLICANT: David Dawson
APPLICANT: Paul Gillis
TITLE OF INVENTION: Methods and Compositions for Inhibiting Angiogenesis
FILE REFERENCE: 0290-23U3
CURRENT APPLICATION NUMBER: US/09/875,223
CURRENT FILING DATE: 2001-06-06
PRIOR APPLICATION NUMBER: US 09/122,079
PRIOR FILING DATE: 1998-07-23
PRIOR APPLICATION NUMBER: PCT/US98/15228
PRIOR FILING DATE: 1998-07-23
PRIOR APPLICATION NUMBER: US 08/899,304
PRIOR FILING DATE: 1997-07-23
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 22484
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Unsure
LOCATION: 1...22484
OTHER INFORMATION: "n" means either a, c, t, or g
US-09-875-223-2

Query Match 4.5%; Score 45; DB 4; Length 22484;
Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 TCCCAAGTCTGGGATTACAGGATGAGCCACCGCGCCGCC 113
|||||
DB 17186 TCCCAAGTCTGGGATTACAGGATGAGCCACCGCGCCGCC 17230

RESULT 10

US-09-875-114-2
; Sequence 2, Application US/09875114
; Patent No. 6670333
; GENERAL INFORMATION:
; APPLICANT: No. 6670333 Western University
; APPLICANT: No. 6670333 Bouck
; APPLICANT: David Dawson
; APPLICANT: Paul Willis
; TITLE OF INVENTION: Methods and Compositions for Inhibiting Angiogenesis
; FILE REFERENCE: 0290-2302
; CURRENT APPLICATION NUMBER: US/09/875,114
; CURRENT FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 09/122,079
; PRIOR FILING DATE: 1998-07-23
; PRIOR APPLICATION NUMBER: PCT/US98/15228
; PRIOR FILING DATE: 1998-07-23
; PRIOR APPLICATION NUMBER: US 08/899,304
; PRIOR FILING DATE: 1997-07-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 2
; LENGTH: 22484
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Unsure
; LOCATION: 1...22484
; OTHER INFORMATION: "n" means either a, c, t, or g
US-09-875-114-2

Query Match 4.5%; Score 45; DB 4; Length 22484;
Best Local Similarity 100.0%; Pred. No. 2.5e-10;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCCGCC 113
Db 17186 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCCGCC 17230

RESULT 11
US-09-984-890-3/c
; Sequence 3, Application US/09984890
; Patent No. 6492156
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001306
; CURRENT APPLICATION NUMBER: US/09/984,890
; CURRENT FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 75395
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(75395)
; OTHER INFORMATION: n = A,T,C or G
US-09-984-890-3

Query Match 4.5%; Score 45; DB 4; Length 75395;
Best Local Similarity 100.0%; Pred. No. 2.3e-10;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCCGCC 113
Db 12402 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCCGCC 12358

RESULT 12

US-09-621-976-11863/c
; Sequence 11863, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET 054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 11863
; LENGTH: 164
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-11863

Query Match 4.4%; Score 44; DB 4; Length 164;
Best Local Similarity 100.0%; Pred. No. 9.2e-10;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 69 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCCGCC 112
Db 59 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCCGCC 16

RESULT 13
US-09-621-976-1944/c
; Sequence 1944, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET 054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 1944
; LENGTH: 281
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 34..195
; NAME/KEY: sig_peptide
; LOCATION: 34..189
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 3.7999995231628
; OTHER INFORMATION: seq LGAVAGACSPSCS/GG
US-09-621-976-1944

Query Match 4.4%; Score 44; DB 4; Length 281;
Best Local Similarity 100.0%; Pred. No. 8.8e-10;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 69 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCCGCC 112
Db 59 TCCCAAAGTGTGGATTACAGGATGAGCCACCGCGCCGCC 16

RESULT 14
US-09-621-976-1939/c
; Sequence 1939, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: ESTs and Encoded Human Proteins.

; FILE REFERENCE: GENSET.054PR2

; CURRENT APPLICATION NUMBER: US/09/621,976

; CURRENT FILING DATE: 2000-07-21

; NUMBER OF SEQ ID NOS: 19335

; SOFTWARE: Patent.pm

; SEQ ID NO 1939

; LENGTH: 294

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 34..195

; NAME/KEY: sig_peptide

; LOCATION: 34..189

; OTHER INFORMATION: Von Heijne matrix

; OTHER INFORMATION: score 3.7999995231628

; OTHER INFORMATION: seq LGAVAGACSPSCS/GG

US-09-621-976-1939

Query Match

Best Local Similarity 4.4%; Score 44; DB 4; Length 294;

Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 69 TCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCCGCC 112

Db 59 TCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCCGCC 16

RESULT 15

US-09-621-976-1940/c

; Sequence 1940, Application US/09621976

; Patent No. 6639063

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Jobert, S.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: ESTs and Encoded Human Proteins.

; FILE REFERENCE: GENSET.054PR2

; CURRENT APPLICATION NUMBER: US/09/621,976

; CURRENT FILING DATE: 2000-07-21

; NUMBER OF SEQ ID NOS: 19335

; SOFTWARE: Patent.pm

; SEQ ID NO 1940

; LENGTH: 294

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 34..195

; NAME/KEY: sig_peptide

; LOCATION: 34..189

; OTHER INFORMATION: Von Heijne matrix

; OTHER INFORMATION: score 3.7999995231628

; OTHER INFORMATION: seq LGAVAGACSPSCS/GG

US-09-621-976-1940

Query Match

Best Local Similarity 4.4%; Score 44; DB 4; Length 294;

Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 69 TCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCCGCC 112

Db 59 TCCCAAAGTCTGGGATTACAGGCATGAGCCACCGCGCCGCC 16

Search completed: July 10, 2004, 15:30:57

Job time : 103 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: July 13, 2004, 08:38:15 ; Search time 375 Seconds
(without alignments)
894.954 Million cell updates/sec

Title: US-09-972-032-2
Perfect score: 79
Sequence: 1 MCGRRVRSAGCGFADAHWT.....SAGLTVRDRPQLGELCMGRG 79

Scoring table:
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Word size: 1
Total number of hits satisfying chosen parameters: 6744130

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09972032/runat_06072004_121533_8862/app_query.fasta_1.263
-DB=N_Geneseq_29Jan04 -OPMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=i -END=s-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORF=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09972032@cgn_1.1_470 @runat_06072004_121533_8862 -NCPU=6 -ICPU=3
-MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FCGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	79	100.0	990	ABL60606	ABL60606 Human ERC
C 2	26	32.9	597	AAS69040	Aas69040 DNA encod
C 3	9	11.4	1402	AAV36085	AAV36085 DNA encod
C 4	8	10.1	379	AAF64670	AAF64670 Novel hum
C 5	8	10.1	483	ABZ37949	Abz37949 N. gonorr
C 6	8	10.1	600	AA65458	Aa65458 Porcine B
C 7	8	10.1	720	ACA23355	Aca23355 Prokaryot
C 8	8	10.1	1015	AAF12681	Aaf12681 Aspergill

C 9	8	10.1	1035	7	ACA25574	Aca25574 Prokaryot
C 10	8	10.1	1037	4	AAH48630	Aah48630 Human MLP
C 11	8	10.1	1230	6	ABQ73845	Abq73845 Rhizobium
C 12	8	10.1	1239	7	ACA25997	Aca25997 Prokaryot
C 13	8	10.1	1266	6	ABQ73844	Abq73844 Yersinia
C 14	8	10.1	1349	2	AAQ06668	Aaq06668 Mycomacte
C 15	8	10.1	1394	6	ABQ79295	Abq79295 Mycobacte
C 16	8	10.1	1396	6	ABK52935	Abk52935 AIDS/hepa
C 17	8	10.1	1437	4	AAQ10372	Aaq10372 Tuberculi
C 18	8	10.1	1437	4	AAS53940	Aas53940 Klebsell
C 19	8	10.1	1536	7	ACA35721	Aca35721 Prokaryot
C 20	8	10.1	1680	4	ABL26019	AbL26019 Drosophi
C 21	8	10.1	1845	7	ACC43487	Acc43487 Nucleotid
C 22	8	10.1	1914	6	ABK51971	Abk51971 Corn cDNA
C 23	8	10.1	1914	7	ABX93207	Abx93207 cDNA enco
C 24	8	10.1	2046	7	ACA28541	Aca28541 Prokaryot
C 25	8	10.1	2169	7	ADA53246	Ada53246 Human cod
C 26	8	10.1	2265	4	ABL09455	AbL09455 Drosophi
C 27	8	10.1	2391	4	AAK77091	Aak77091 Human imm
C 28	8	10.1	2728	6	ABQ99435	Abq99435 Human cod
C 29	8	10.1	2883	4	ABL25994	AbL25994 Drosophi
C 30	8	10.1	2929	6	ABK81822	Abk81822 DNA repre
C 31	8	10.1	3680	4	ABL26018	AbL26018 Drosophi
C 32	8	10.1	3752	3	AAA65427	Aaa65427 Porcine B
C 33	8	10.1	5877	6	ABS78681	Abs78681 Kitasatos
C 34	8	10.1	9857	4	ABL09454	AbL09454 Drosophi
C 35	8	10.1	21034	2	AAV62154	Aav62154 HSV-2 str
C 36	8	10.1	22428	4	AA541759	Aa541759 Genomic s
C 37	8	10.1	22428	7	ABZ67767	Abz67767 Human sec
C 38	8	10.1	22428	7	ADA98730	Ada98730 Human sec
C 39	8	10.1	26338	2	AAV62134	Aav62134 HSV-2 str
C 40	8	10.1	36955	6	ABV73608	Abv73608 S. albulu
C 41	8	10.1	49999	2	AAZ23302	Aaz23302 Human LOB
C 42	8	10.1	92934	3	AA81473	Aa81473 N. mening
C 43	8	10.1	92934	3	AA81473	Aa81473 N. mening
C 44	8	10.1	110000	3	AA81489_7	Continuation (8 of
C 45	8	10.1	110000	4	AA199682_09	Continuation (10 of

ALIGNMENTS

RESULT 1
ABL60606
ID ABL60606 standard; cDNA; 990 BP.

XX ABL60606;
XX
XX 27-AUG-2002 (first entry)
XX
XX Human ERCoA3 protein encoding cDNA.
XX
XX Estrogen Receptor Coregulator 3; ERCoA3; tamoxifen; cancer;
XX osteoporosis; cytostatic; osteopathic; human; gene; ss.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 203..442
XX FT /*tag= a
XX FT /product= "ERCoA3 protein"
XX
XX WO200228352-A2.

PD 11-APR-2002.
XX
XX 05-OCT-2001; 2001WO-US031271.
XX
XX 05-OCT-2000; 2000US-0238190P.
XX
XX (UYCA-) UNIV CASE WESTERN RESERVE.
XX
XX Montano M, Sutton A;
XX

DR WPI: 2002-454492/48.
 DR P-PSDB; ABB08035.
 XX
 PT New polypeptide, that is a functional equivalent of ERCoA3 (Estrogen
 PT Receptor Coregulator 3), is useful in inhibiting or reducing tamoxifen or
 PT estrogen-induced proliferation of cancer cells and in treating
 PT osteoporosis.
 XX
 PS Claim 5; Fig 1; 39pp; English.
 XX
 CC The invention relates to a ERCoA3 (Estrogen Receptor Coregulator 3)
 CC protein and encoding polynucleotides. ERCoA3 can be used to inhibit or
 CC reduce tamoxifen or estrogen induced proliferation of cancer cells, by
 CC reduced activity of ERCoA3, and for detecting cancer cells that are
 CC tamoxifen resistant, or to treat osteoporosis, by increasing levels of
 CC ERCoA3 in cells. The encoding polynucleotide can be used to inhibit
 CC translation of a mRNA encoding ERCoA3. ERCoA3 acts as a coregulator
 CC protein and can bind to the estrogen receptor to activate a molecular or
 CC cellular response in the cell. The present sequence represents the human
 CC ERCoA3 encoding cDNA
 XX
 SQ Sequence 990 BP; 189 A; 281 C; 317 G; 203 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 4.88e-66 Length: 990
 Score: 79.00 Matches: 79
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-09-972-032-2 (1-79) x ABL60606 (1-990)
 QY 1 MetCysGlyArgProArgArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThr 20
 Db 203 ATGTGTGGGAGACCGCGTCGCGTAAGCGCTGGATGTGGCTTCGCTGATGCACATTGGACC 262
 QY 21 GlyLeuTrpThrGlyLeuGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGln 40
 Db 263 GGGCTCTGGACTGGCTAGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 322
 QY 41 SerProThrProAspCysAlaSerArgTrpProArgSerAlaSerArgTrpProTrpSer 60
 Db 323 TCGCCGACCCCGACCTCGCGCTCCCGTGGCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 382
 QY 61 AlaGlyLeuThrValArgAspArgProGlnLeuGlyGlnLeuCysMetClyArgGly 79
 Db 383 GCAGGTCTTACCGTCCGAGATCGTCCGCAACTGGGCGAGCTGTGCATGGGGCGGTGGC 439
 RESULT 2
 AAS69040/c
 ID AAS69040 standard; cDNA; 597 BP.
 XX
 AC AAS69040;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #4844.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 OS Homo sapiens.
 XX
 PN WC200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US008631.
 XX
 PR 31-MAR-2000; 2000US-00540217.
 PR 23-AUG-2000; 2000US-00649167.
 XX

PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG04853.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 1; SEQ ID NO 4844; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
 CC sequences. (I) is useful as hybridisation probes, polymerase chain
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
 CC and in recombinant production of (II). The polynucleotides are also used
 CC in diagnostics as expressed sequence tags for identifying expressed
 CC genes. (I) is useful in gene therapy techniques to restore normal
 CC activity of (II) or to treat disease states involving (II). (II) is
 CC useful for generating antibodies against it, detecting or quantitating a
 CC polypeptide in tissue, as molecular weight markers and as a food
 CC supplement. (II) and its binding partners are useful in medical imaging
 CC of sites expressing (II). (I) and (II) are useful for treating disorders
 CC involving aberrant protein expression or biological activity. The
 CC polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
 CC coding sequences of the invention. Note: The sequence data for this
 CC patent did not appear in the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 597 BP; 97 A; 217 C; 196 G; 87 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.01e-15 Length: 597
 Score: 26.00 Matches: 26
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 32.91% Indels: 0
 DB: 5 Gaps: 0
 US-09-972-032-2 (1-79) x AAS69040 (1-597)
 QY 1 MetCysGlyArgProArgArgValSerAlaGlyCysGlyPheAlaAspAlaHisTrpThr 20
 Db 522 ATGTGTGGGAGACCGCGTCGCGTAAGCGCTGGATGTGGCTTCGCTGATGCACATTGGACC 463
 QY 21 GlyLeuTrpThrGlyLeu 26
 Db 462 GGGCTCTGGACTGGGCTG 445
 RESULT 3
 AAV36085/c
 ID AAV36085 standard; DNA; 1402 BP.
 XX
 AC AAV36085;
 XX
 DT 17-OCT-2003 (revised)
 DT 03-SEP-1998 (first entry)
 XX
 DE DNA encoding a 36K antigen of Mycobacterium paratuberculosis.
 XX
 KW Clone pMtb #48; 36K antigen; detection; Mycobacteria; diagnosis;
 KW infection; Crohn's disease; sarcoidosis; serological diagnosis;
 KW vaccine development; ss.
 XX
 OS Mycobacterium avium subsp. paratuberculosis.
 XX

PN US5776692-A.
 XX
 PD 07-JUL-1998.
 XX
 PF 23-MAY-1995; 95US-00447965.
 XX
 PR 23-MAY-1995; 95US-00447965.
 XX
 PA (BAVU) BAYLOR COLLEGE MEDICINE.
 XX
 PI Graham DY, El-Zaatari FAK, Naser S;
 XX
 DR WPI; 1998-398024/34.
 DR P-PSDB; AAW60723.
 XX
 PT Recombinant clone encoding new Mycobacterium paratuberculosis protein -
 PT containing nucleic acid useful as genus-specific hybridisation probe for
 PT detecting mycobacteria.
 XX
 PS Claim 2; Fig 5; 12pp; English.
 XX
 CC The present sequence represents the BamHI-DNA insert of the recombinant
 CC clone pMpb #48 and encodes a 36K antigen of Mycobacterium
 CC paratuberculosis. The DNA sequence and its fragments are useful as
 CC hybridisation probes and amplification primers for detecting nucleic acid
 CC specific for the genus Mycobacterium, particularly for diagnosis of
 CC infection, especially by M. paratuberculosis which has been implicated in
 CC Crohn's disease and sarcoidosis. The 36K protein is useful in serological
 CC diagnosis and for vaccine development. (Updated on 17-OCT-2003 to
 CC standardise OS field)
 XX
 SQ Sequence 1402 BP; 171 A; 490 C; 552 G; 185 T; 0 U; 4 Other;
 XX
 Alignment Scores:
 Pred. No.: 35-7 Length: 1402
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 11.39% Indels: 0
 DB: 2 Gaps: 0

US-09-972-032-2 (1-79) x AAV36085 (1-1402)
 QY 47 AlaserArgTrpProArgSerAlaser 55
 DB 1260 GCGTCCCGTTGGCCCGAGATCGGCAGT 1234

RESULT 4
 AAF64670/C
 ID AAF64670 standard; cDNA; 379 BP.
 XX
 AC AAF64670;
 XX
 DT 09-APR-2001 (first entry)
 XX
 DE Novel human polynucleotide, SEQ ID NO: 426.
 XX
 KW Human; cytostatic; gene therapy; colon cancer; prostate cancer;
 KW breast cancer; lung cancer; cancer detection; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO2001025568-A2.
 XX
 PD 11-JAN-2001.
 XX
 PF 30-JUN-2000; 2000WO-US018374.
 XX
 PR 02-JUL-1999; 99US-0142310P.
 PR 02-JUL-1999; 99US-0142311P.
 XX
 PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.

Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J;
 Kassar A, Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G;
 Drmanac R, Crkenjakov R, Drmanac S, Dickson M, Labat I;
 Leshkowitz D, Kita D, Garcia V, Jones LW, Strache-Crain B;
 WPI; 2001-091805/10.
 XX
 XX Library of polynucleotides for diagnosing a cancerous state of a
 PT mammalian cell and detecting cancer, particularly of the colon or
 PT prostate, comprises 3351 human polynucleotide sequences.
 XX
 PS Claim 9; Page 605-606; 1046pp; English.
 XX
 CC The present sequence is one of 3351 sequences in a library of human
 CC polynucleotides. The library is used to detect differentially expressed
 CC genes correlated with a cancerous state of a mammalian cell and can
 CC detect colon, prostate, breast and lung cancer. The library can be used
 CC to produce probes for detection of mRNA and to produce additional copies
 CC of the polynucleotides. The probes can be used for chromosome mapping of
 CC the polynucleotide and for detection of transcription levels. Ribozymes
 CC or antisense oligonucleotides can be generated. The polynucleotides and
 CC their gene products are used as genetic or biochemical markers (e.g. in
 CC blood or tissues) that will detect the earliest changes along the
 CC carcinogenesis pathway and/or monitor the efficacy of therapies and
 CC preventive interventions. The polynucleotides, polypeptides and
 CC antibodies against them can be used in pharmaceutical compositions to
 CC treat the cancers and proliferative disorders such as neoplasia,
 CC dysplasia and hyperplasia
 XX
 SQ Sequence 379 BP; 79 A; 108 C; 110 G; 82 T; 0 U; 0 Other;
 XX
 Alignment Scores:
 Pred. No.: 95.4 Length: 379
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.13% Indels: 0
 DB: 5 Gaps: 0

US-09-972-032-2 (1-79) x AAF64670 (1-379)
 QY 35 GlyProGluGlyGlnAlaserPro 42
 DB 30 GGGCCTGAGGACAGGCCAGTCCT 7

RESULT 5
 ABZ37949/C
 ID ABZ37949 standard; DNA; 483 BP.
 XX
 AC ABZ37949;
 XX
 DT 07-MAR-2003 (first entry)
 XX
 DE N. gonorrhoeae nucleotide sequence SEQ ID 487.
 XX
 KW Antibacterial; infection; vaccine; gene therapy; gene; ds.
 XX
 OS Neisseria gonorrhoeae.
 XX
 PN WO200279243-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 12-FEB-2002; 2002WO-IB002069.
 XX
 PR 12-FEB-2001; 2001GB-00003424.
 XX
 PA (CHIR-) CHIRON SPA.
 XX
 PI Fontana MR, Pizza M, Masignani V, Monaci E;
 XX
 DR WPI; 2003-058415/05.

DR P-PSDB; ABP76979.
 XX New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a
 PT medicament for treating or preventing *N. gonorrhoeae* infection.
 XX
 PS Disclosure; Page 224; 815pp; English.
 XX
 CC The present invention relates to proteins from *Neisseria gonorrhoeae*.
 CC Also disclosed are the nucleic acid molecules encoding the proteins and
 CC antibodies that specifically bind to the proteins. The composition
 CC comprising the protein, nucleic acid or antibody is useful for the
 CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*
 CC infection, this may be in the form of a vaccine or gene therapy.
 CC Sequences given in records AB237706-AB242016 represent nucleic acid
 CC molecules of the invention
 XX
 SQ Sequence 483 BP; 110 A; 172 C; 124 G; 77 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 119 Length: 483
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.13% Indels: 0
 DB: 7 Gaps: 0
 US-09-972-032-2 (1-79) x AB237949 (1-483)
 QY 7 ArgValSerAlaGlyCysGlyPhe 14
 |||||
 DB 442 CGTGATCCAGCAGATGCGTTT 419
 RESULT 6
 AAA65458/c
 ID AAA65458 standard; DNA; 600 BP.
 XX
 AC AAA65458;
 XX
 DT 10-NOV-2000 (first entry)
 XX
 DE Porcine BAC-PIGF2-1 contig 49.
 XX
 KW Porcine; pig; wild boar; quantitative trait locus; QTL; chromosome 2;
 KW mapping; 2p1.7; select breeding; genotype; phenotype; muscle mass;
 KW fat deposition; IGF2; insulin-like growth factor 2; ds.
 XX
 OS Sus scrofa.
 XX
 PN WO200036143-A2.
 XX
 PD 22-JUN-2000.
 XX
 PF 16-DEC-1999; 99WO-EP010209.
 XX
 PR 16-DEC-1998; 98EP-00204291.
 XX
 PA (UJLI-) UNIV LIEGE.
 PA (MELI-) MELICK HB.
 PA (SEGH-) SEGHERSGENTEC NV.
 XX
 PI Andersson L, Georges M, Spincemaille G;
 XX
 DR WPI; 2000-431612/37.
 XX
 PT Selecting a domestic animal for having desired genotypic properties
 PT comprises testing the animal for the presence of a parentally imprinted
 PT quantitative trait locus which is related to muscle mass and/or fat
 PT deposition.
 XX
 PS Example 3; Fig 6; 107pp; English.
 XX
 CC The present invention describes a method (M1) for selecting a domestic
 CC animal for having desired genotypic properties. The method comprises

CC testing the animal for the presence of a parentally imprinted
 CC quantitative trait locus (QTL). The pig QTL is located at chromosome 2,
 CC mapping at around position 2p1.7. Also described are: (1) an isolated
 CC and/or recombinant nucleic acid (N1) comprising a parentally imprinted
 CC QTL or its functional fragment; (2) an isolated and/or recombinant
 CC nucleic acid (N2) comprising a synthetic parentally imprinted QTL derived
 CC from at least one chromosome or its functional fragment; (3) an animal
 CC such as pig selected for having desired genotypic or potential phenotypic
 CC properties; (4) a transgenic animal comprising N1 or N2; and (5) sperm or
 CC an embryo derived from the animal of (3) or (4). N1 or its fragment is
 CC useful for selecting an animal destined for slaughter or a breeding
 CC animal having desired genotypic or potential phenotypic properties. The
 CC properties are related to muscle mass and/or fat deposition. The sperm or
 CC an embryo are useful in breeding animals destined for slaughter. AAA65418
 CC to AAA65524 represent contigs 1 to 10 and 19 to 115 which were isolated
 CC from porcine BAC-PIGF2-1 which contains the INS and IGF2 (insulin-like
 CC growth factor) genes. These sequences were used in an example from the
 CC present invention for generating a reference sequence of IGF2 and
 CC flanking loci in the pig
 XX
 SQ Sequence 600 BP; 102 A; 162 C; 228 G; 108 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 146 Length: 600
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.13% Indels: 0
 DB: 3 Gaps: 0
 US-09-972-032-2 (1-79) x AAA65458 (1-600)
 QY 38 GlyGlnAlaSerProThrProAsp 45
 |||||
 DB 365 GGCCAAGCGTCACCGACCCCTGAC 342
 RESULT 7
 ACA23355/c
 ID ACA23355 standard; DNA; 720 BP.
 XX
 AC ACA23355;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Prokaryotic essential gene #5012.
 XX
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.
 XX
 OS *Borrelia cepacia*.
 XX
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SEP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR P-PSDB; ABU19485.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.
 XX Claim 14; SEQ ID NO 11225; 1766pp; English.
 XX
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: the sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 720 BP; 127 A; 259 C; 233 G; 101 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 173 Length: 720
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.13% Indels: 0
 DB: 7 Gaps: 0

US-09-972-032-2 (1-79) x ACA23355 (1-720)

Qy 47 AlaSerArgTrpProArgSerAla 54
 |||||
 Db 107 GCGAGCGCGTGGCGGCTTCGGCC 84
 |||||

RESULT 8
 AAF12681
 ID AAF12681 standard; cDNA; 1015 BP.
 AC AAF12681;
 XX
 XX 13-MAR-2001 (first entry)
 DE
 DE Aspergillus oryzae EST SEQ ID NO:5204.
 KW Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX Aspergillus oryzae.
 OS
 XX WO2000056762-A2.
 PN
 XX 28-SEP-2000.
 PD

XX 22-MAR-2000; 2000WO-US007781.
 XX
 XX 22-MAR-1999; 99US-00273623.
 PR
 XX (NOVO) NOVO NORDISK BIOTECH INC.
 PA (NOVO) NOVO NORDISK AS.
 PA
 XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
 PI WPI; 2000-594572/56.
 XX
 XX Monitoring differential expression of genes in filamentous fungal cells
 PT uses fluorescence-labeled nucleic acids isolated from the cells and a
 PT substrate of expressed sequence tags.
 XX
 XX Claim 88; Page 2176-2177; 3161pp; English.
 XX
 CC The present invention describes a method for monitoring differential
 CC expression of genes in a first filamentous fungal (FF) cell relative to
 CC expression of the same genes in one or more second filamentous fungal
 CC cells. The method uses fluorescence-labeled nucleic acids isolated from
 CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
 CC are used in the methods for monitoring differential expression of genes
 CC in a first filamentous fungal (FF) cell relative to expression of the
 CC same genes in one or more second filamentous fungal cells. Monitoring the
 CC global expression of genes from FF cells allows the production potential
 CC of the microorganisms to be improved. New genes may be discovered,
 CC possible functions of unknown open reading frames can be identified and
 CC gene copy number variation and stability can be monitored. The expression
 CC of genes can be used to study how FF cells adapt to changes in culture
 CC conditions, environmental stress, spore morphogenesis, recombination,
 CC metabolic or catabolic pathway engineering. Using ESTs provides several
 CC advantages over genomic or random cDNA clones including elimination of
 CC redundancy as one spot on an array equals one gene or open reading frame,
 CC and organisation of the microarrays based on function of the gene
 CC products to facilitate analysis of the results. AAF07478 to AAF11247
 CC represents ESTs from *Fusarium venenatum*; AAF11248 to AAF11853 represents
 CC ESTs from *Aspergillus niger*; AAF11854 to AAF14878 represents ESTs from
 CC *Aspergillus oryzae*; and AAF14879 to AAF15337 represents ESTs from
 CC *Trichoderma reesei*, which are all specifically claimed in the present
 CC invention
 XX
 SQ Sequence 1015 BP; 222 A; 313 C; 249 G; 230 T; 0 U; 1 Other;

Alignment Scores:
 Pred. No.: 238 Length: 1015
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.13% Indels: 0
 DB: 3 Gaps: 0

US-09-972-032-2 (1-79) x AAF12681 (1-1015)

Qy 49 ArgTrpProArgSerAlaSerArg 56
 |||||
 Db 602 CGCTGGCCGCGAGCGCTCCCGG 625
 |||||

RESULT 9
 ACA25574/c
 ID ACA25574 standard; DNA; 1035 BP.
 XX
 AC ACA25574;
 XX
 XX 19-JUN-2003 (first entry)
 DT
 XX Prokaryotic essential gene #7231.
 DE
 XX Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW Antisense; gene.
 XX
 XX Burkholderia fungorum.
 OS

XX WO200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX P-PSDB; ABU21704.
XX WPI; 2003-029926/02.
XX DR P-PSDB; ABU21704.
XX New antisense nucleic acids, useful for identifying proteins or screening
PI for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 14; SEQ ID NO 13444; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 5213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 1035 BP; 235 A; 290 C; 306 G; 204 T; 0 U; 0 Other;

Alignment Scores: 242 Length: 1035
Pred. No.: 8.00 Matches: 8
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 10.13% Gaps: 0
DB: 7

US-09-972-032-2 (1-79) x ACA25574 (1-1035)

QY 48 SerArgTrpProArgSerAlaSer 55
|||||

Db 836 TCGAGATGCCCGAGGTCTGCCAGT 813
RESULT 10
AAH48630/c
ID AAH48630 standard; DNA; 1037 BP.
XX AAH48630;
XX 21-SEP-2001 (first entry)
XX Human MLP promoter fragment.
XX MLP; human; mutation; muscle-specific promoter; cardiovascular disease;
KW dilative cardiomyopathy; cardiac; gene therapy; myocardial disease;
KW sarcomer; dystrophin; cardiac actin; hypertrophic cardiomyopathy;
KW long QT syndrome; chromosome 1p15.1; promoter; ds.
XX Homo sapiens.
XX WO200157208-A2.
XX 09-AUG-2001.
XX 01-FEB-2001; 2001WO-EP001042.
XX 03-FEB-2000; 2000DE-01004857.
XX (SCHD) SCHERING AG.
XX Knoell R;
XX WPI; 2001-483436/52.
XX New nucleic acid encoding mutant MLP, useful for diagnosis and treatment
PT of myocardial disease, particularly dilative cardiomyopathy.
XX Claim 35; Page 51; 53pp; German.
XX This invention describes a novel nucleic acid (I) encoding an MLP (not
CC defined) which has a 1273 base pair (bp) sequence (1) that includes a
CC mutation at base 10 in exon 2 or the third position of codon 112 in exon
CC 4, is new. The product of the invention has cardiac activity and can be
CC used for gene therapy. (I), and related nucleic acids or probes, are used
CC in diagnosis of and/or screening for myocardial diseases (or
CC predisposition), especially dilative cardiomyopathy. Both specified
CC mutations are associated with development of these diseases. Antibodies
CC (Ab) raised against MCP and other peptides encoded by (I) can be used
CC similarly. Also the regulatory region (III) of the genomic MLP sequence
CC (optionally when incorporated into vectors or cells) is used in gene
CC therapy, specifically for prevention and/or treatment of cardiovascular
CC disease, particularly those which involve a point mutation in a gene
CC encoding sarcomer, dystrophin or cardiac actin, e.g. hypertrophic
CC cardiomyopathy, long QT syndrome and dilative cardiomyopathy. The
CC regulatory region of the MLP gene provides muscle-specific gene
CC expression. This sequence represents a promoter fragment derived from the
CC human MLP described in the invention
XX SQ Sequence 1037 BP; 285 A; 232 C; 232 G; 288 T; 0 U; 0 Other;

Alignment Scores: 242 Length: 1037
Pred. No.: 8.00 Matches: 8
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 10.13% Gaps: 0
DB: 4

US-09-972-032-2 (1-79) x AAH48630 (1-1037)

QY 26 LeuGlyGluGlyGlnGlyGly 33
|||||

Db 910 CTGGGGGAAGGCGAGGAGGAGGC 887

RESULT 11
 ABQ73845
 ID ABQ73845 standard; DNA; 1230 BP.
 XX
 AC ABQ73845;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Rhizobium phytase nucleotide sequence SEQ ID NO:7.
 XX
 KW Rhizobium; phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed; gene;
 KW ds.
 XX
 OS Rhizobium sp.
 XX
 FH Key Location/Qualifiers
 CDS 1..1230
 FT /*tag= a
 FT /EC_number= "3.1.3.8"
 FT /product= "phytase"
 XX
 PN WO200248332-A2.
 XX
 PD 20-JUN-2002.
 XX
 PF 12-DEC-2001; 2001WO-US048774.
 XX
 PR 12-DEC-2000; 2000US-0255090P.
 XX
 PA (DIVE-) DIVERSA CORP.
 XX
 PI Short J, Mathur EJ, Richardson T, Robertson D, Barton N;
 XX
 DR WPI; 2002-583504/62.
 DR P-PSDB; ABP51932.
 XX
 PT Novel recombinant phytase protein and polynucleotide for improving
 PT nutritional value of phytate-containing foodstuff, in animal feed and
 PT feed supplements and to degrade excess phytase from environment or
 PT sample.
 XX
 PS Claim 1; Fig 5G; 208pp; English.
 CC
 CC The present invention describes an isolated phytase protein (I). (I) can
 CC be used for improving the nutritional value of a phytate-containing
 CC foodstuff, by contacting the phytate-containing foodstuff with (I), where
 CC the enzyme catalyses the liberation of inorganic phosphate from the
 CC phytate-containing foodstuff, and so improving the nutritive value of the
 CC contacted foodstuff. The liberation of inorganic phosphate occurs prior
 CC to or after the ingestion of phytate-containing foodstuff by a recipient
 CC organism. Nucleotide sequences (II) encoding (I) can be used for
 CC producing an animal feed, by transforming a plant, plant portion or plant
 CC cell with a nucleic acid expression vector, comprising (II), culturing
 CC the plant, plant portion or plant cell under conditions in which the
 CC phytase protein is expressed and converting the plant, plant portion or
 CC plant cell into a composition suitable for animal feed. The animal is
 CC preferably a monogastric animal or a ruminant. (I) and (II) are useful in
 CC animal feed and feed supplements as well as in treatments to degrade or
 CC remove excess phytate from the environment or a sample. (I) reduces
 CC phytate levels in animal manure and so reduces the phosphate pollution of
 CC the environment. The present sequence encodes a phytase from the present
 CC invention
 XX
 SQ Sequence 1230 BP; 182 A; 434 C; 419 G; 195 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 284 Length: 1230
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.13% Indels: 0
 DB: 6 Gaps: 0

US-09-972-032-2 (1-79) x ABQ73845 (1-1230)
 QY 47 AlaserArgTTPProArgSerAla 54
 Db 725 GCAAGCCGATGCCCGAGTCGGCT 748
 RESULT 12
 ACA25997
 ID ACA25997 standard; DNA; 1239 BP.
 XX
 AC ACA25997;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Prokaryotic essential gene #7654.
 XX
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.
 XX
 OS Burkholderia mallei.
 XX
 PN WO200277183-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 21-MAR-2002; 2002WO-US009107.
 XX
 PR 21-MAR-2001; 2001US-00815242.
 PR 06-SRP-2001; 2001US-00948993.
 PR 25-OCT-2001; 2001US-0342923P.
 PR 08-FEB-2002; 2002US-00072851.
 PR 06-MAR-2002; 2002US-0362699P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX
 DR WPI; 2003-029926/02.
 DR P-PSDB; ABU22127.
 XX
 PT New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX
 PS Claim 14; SEQ ID NO 13867; 1766pp; English.
 CC
 CC The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,

CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 1239 BP; 163 A; 419 C; 480 G; 177 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 286 Length: 1239
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.13% Indels: 0
 DB: 7 Gaps: 0

US-09-972-032-2 (1-79) x ACA25997 (1-1239)

QY 48 SerArgTrpProArgSerAlaSer 55
 DB 95 AGCAGATGCCCGCAGCGCTTCA 118

RESULT 13
 ABQ73844
 ID ABQ73844 standard; DNA; 1266 BP.
 XX
 AC ABQ73844;
 XX
 DT 08-OCT-2002 (first entry)
 XX
 DE Yersinia pestis phytase nucleotide sequence SEQ ID NO:5.
 XX
 KW Yersinia pestis; phytase; enzyme; E.C.3.1.3.8; phytate; foodstuff; feed;
 KW Gene; ds.
 XX
 OS Yersinia pestis.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1266
 FT /*tag= a
 FT /EC number= "3.1.3.8"
 FT /product= "phytase"
 XX
 PN WO200248332-A2.
 XX
 PD 20-JUN-2002.
 XX
 PF 12-DEC-2001; 2001WO-US048774.
 XX
 PR 12-DEC-2000; 2000US-0255090P.
 XX
 PA (DIVE-) DIVERSA CORP.
 XX
 PI Short J, Mathur EJ, Richardson T, Robertson D, Barton N;
 XX
 DR WPI; 2002-583504/62.
 DR P-PSDB; ABP51931.
 XX
 XX Novel recombinant phytase protein and polynucleotide for improving
 PT nutritional value of phytate-containing foodstuff, in animal feed and
 PT feed supplements and to degrade excess phytase from environment or
 PT sample.
 XX
 PS Claim 1; Fig 5E; 208pp; English.
 XX
 CC The present invention describes an isolated phytase protein (I). (I) can
 CC be used for improving the nutritional value of a phytate-containing
 CC foodstuff, by contacting the phytate-containing foodstuff with (I), where
 CC the enzyme catalyses the liberation of inorganic phosphate from the
 CC phytate-containing foodstuff, and so improving the nutritive value of the
 CC contacted foodstuff. The liberation of inorganic phosphate occurs prior
 CC to or after the ingestion of phytate-containing foodstuff by a recipient
 CC organism. Nucleotide sequences (II) encoding (I) can be used for

CC producing an animal feed, by transforming a plant, plant portion or plant
 CC cell with a nucleic acid expression vector, comprising (II), culturing
 CC the plant, plant portion or plant cell under conditions in which the
 CC phytase protein is expressed and converting the plant, plant portion or
 CC plant cell into a composition suitable for animal feed. The animal is
 CC preferably a monogastric animal or a ruminant. (I) and (II) are useful in
 CC animal feed and feed supplements as well as in treatments to degrade or
 CC remove excess phytate from the environment or a sample. (I) reduces
 CC phytate levels in animal manure and so reduces the phosphate pollution of
 CC the environment. The present sequence encodes a phytase from the present
 CC invention
 XX
 SQ Sequence 1266 BP; 194 A; 444 C; 426 G; 202 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 292 Length: 1266
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.13% Indels: 0
 DB: 6 Gaps: 0

US-09-972-032-2 (1-79) x ABQ73844 (1-1266)

QY 47 AlaSerArgTrpProArgSerAla 54
 DB 761 GCAGCCGATGCCGAGGTCGGCT 784

RESULT 14
 AAQ06668
 ID AAQ06668 standard; DNA; 1349 BP.
 XX
 AC AAQ06668;
 XX
 DT 01-MAR-1991 (first entry)
 XX
 DE Mycomacterium derived alpha-antigen gene.
 XX
 KW Vaccine; tuberculosis; HIV-1; ds.
 XX
 OS Mycobacterium kansasii.
 XX
 FH Key Location/Qualifiers
 FT CDS 270..1244
 FT /*tag= a
 FT sig_peptide 270..389
 FT /*tag= b
 XX
 PN EP400973-A.
 XX
 PD 05-DEC-1990.
 XX
 PF 30-MAY-1990; 90EP-00305849.
 XX
 PR 31-MAY-1989; 89JP-00135855.
 PR 16-MAR-1990; 90JP-00064310.
 XX
 PA (AJIN) AJINOMOTO KK.
 XX
 XX Matsuo K, Yamaguchi R, Yamazaki A, Yamada T;
 PI WPI; 1990-363461/49.
 XX P-PSDB; AAR08099.
 DR
 XX New mycobacterial secretory vector - used to transform host cells, and
 PT vaccine comprising transformant.
 XX
 PS Disclosure; Fig 1; 19pp; English.
 XX
 CC The sequence may be used to produce a recombinant protein comprising
 CC alpha-antigen and B-cell epitope of the HIV-1 gag antigen p17. The
 CC product is useful in vaccine production eg. tuberculosis. A live vaccine
 CC such as M.bovis BCG or M.smegmatis has a low toxicity and long lasting

CC activity
 SQ Sequence 1349 BP; 232 A; 434 C; 438 G; 245 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 309 Length: 1349
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.13% Indels: 0
 DB: 2 Gaps: 0

US-09-972-032-2 (1-79) x AAQ06668 (1-1349)

Qy 51 ProArgSerAlaSerArgTrpPro 58
 Db 751 CCGCGTCCGCATCTCGATGCCG 774

RESULT 15

ABQ79295
 ID ABQ79295 standard; cDNA; 1394 BP.
 XX
 AC ABQ79295;
 XX
 DT 15-NOV-2002. (first entry)
 XX
 DE Mycobacterium-originated alpha antigen encoding sequence.
 XX
 KW Antiallergic; Dermatological; Antiasthmatic; Antiinflammatory;
 KW Ophthalmological; Mycobacterium-originated alpha antigen;
 KW allergic disease; atopic dermatitis; asthma; allergic rhinitis;
 KW allergic conjunctivitis; interleukin-4; Th2-type cytokine;
 KW IGE production; histamine; eosinophilic infiltration; gene; ss.
 XX

OS Mycobacterium kansasii.

XX WO20026055-A1.

XX 29-AUG-2002.

XX 20-FEB-2002; 2002WO-JP001459.

XX 20-FEB-2001; 2001JP-00043291.

XX (PRIM-) PRIMMUNE CORP INC.

XX Yasutomi Y, Mizutani H;

XX WPI; 2002-667038/71.

XX Medicinal use of Mycobacterium-originated alpha-antigen or its gene in
 PT treating allergic diseases e.g. atopic dermatitis, asthma, allergic
 PT rhinitis and allergic conjunctivitis.

XX Disclosure; Page 26-27; 34pp; Japanese.

XX This invention relates to pharmaceutical compositions for preventing or
 CC treating allergic diseases containing Mycobacterium-originated alpha
 CC antigens. The compositions are antiallergic, dermatological,
 CC antiasthmatic, antiinflammatory and ophthalmological. The antigen and its
 CC encoded gene are for producing drug compositions in treating allergic
 CC diseases e.g. atopic dermatitis, asthma, allergic rhinitis and allergic
 CC conjunctivitis. The drug compositions are made from the protein that can
 CC inhibit production of interleukin-4, ameliorate Th2-type cytokine-
 CC predominate immune state and relieve symptoms of allergic diseases such
 CC as IGE production, histamine release and eosinophilic infiltration. The
 CC present sequence represents the encoding sequence for a mycobacterium-
 CC originated alpha antigen. This encoding sequence does not encode the
 CC protein featured (ABB8362) despite being said to do so in the
 CC specification

XX Sequence 1394 BP; 241 A; 452 C; 449 G; 252 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 319 Length: 1394
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.13% Indels: 0
 DB: 6 Gaps: 0

US-09-972-032-2 (1-79) x ABQ79295 (1-1394)

Qy 51 ProArgSerAlaSerArgTrpPro 58

Db 751 CCGCGTCCGCATCTCGATGCCG 774

Search completed: July 13, 2004, 15:05:28
 Job time : 394 secs

This Page Blank (uspo)